The Facts behind the Figures:

The Validation of The Older Persons and Informal Caregivers Minimum Data Set

Jennifer E. Lutomski

The Facts behind the Figures: The Validation of The Older Persons and Informal Caregivers Minimum Data Set

The work described in this thesis was funded by the Dutch National Care for the Elderly Programme, coordinated and sponsored by ZonMw (The Organization for Health Research and Development, The Netherlands). Grant reference number: 310300002

Financial support for the publication of this thesis was kindly provided by the Department of Geriatric Medicine, Radboud University Medical Center.

ISBN
978-90-9029628-9

Design/lay-out
Promotie In Zicht, Arnhem

Print
Ipskamp Printing, Enschede

The Facts behind the Figures:

The Validation of The Older Persons and Informal Caregivers Minimum Data Set

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 31 mei 2016
om 10.00 uur precies

door

Jennifer Elizabeth Lutomski
geboren op 15 februari 1982
te New Haven, Connecticut, USA

Promotor
Prof. dr. M.G.M. Olde Rikkert

Copromotor
Dr. R.J.F. Melis

Manuscriptcommissie
Prof. dr. J. Gussekloo (LUMC)
Dr. T.R.J. Schermer
Prof. dr. A.L.M. Verbeek

The Facts behind the Figures:

The Validation of The Older Persons and Informal Caregivers Minimum Data Set

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus,
according to the decision of the Council of Deans
to be defended in public on Tuesday, May 31, 2016
at 10:00am

by

Jennifer Elizabeth Lutomski
born on February 15, 1982
in New Haven, Connecticut, USA

Supervisor
Prof. M.G.M. Olde Rikkert

Co-supervisor
Dr. R.J.F. Melis

Manuscript Committee
Prof. J. Gussekloo
Dr. T.R.J. Schermer
Prof. A.L.M. Verbeek

This thesis is dedicated in loving memory of
Joseph Gee and Charles Lutomski, Sr.

Joe, the little red-haired girl did it.
Grandpa, across the auditorium, I hear you shout, "That's my granddaughter!"

Table of Contents

1

General Introduction

1.1. Why create a minimal dataset?

The ever-evolving discipline of public health has prompted the development of myriad instruments to measure health and wellbeing. Whereas some instruments have become cornerstones in international research, many continue to be used on a more local level. Consequently, health monitoring systems, clinical audits, and medical research studies collect outcome data which are not always directly comparable or even optimal in the chosen setting. Disconcertingly, some instruments purporting to measure the same outcome may be operationalized differently further inhibiting comparative analyses. This fragmentation in data collection impedes the efficient and effective evaluation of interventions as well as the development of clear health policy guidelines and recommendations.

To counter this issue, there has been a movement towards developing minimum datasets with standardized measures to assess changes in functional status, health status, and quality of care.[1,2] Core outcomes are typically chosen to capture data relevant for clinical, research, and policy purposes. In geriatric research, one of the most well-known minimum datasets is the Resident Assessment Instrument Minimum Data Set (simply referred to as MDS).[3] Stemming from the U.S. 1987 Nursing Home Reform Act, the MDS was developed to combat abuse, neglect, and substandard care in nursing homes across the United States. Using this instrument, standardized information on physical and mental wellbeing is collected from nursing home residents upon admission and at periodic intervals. These data are then centrally pooled to allow for independent performance evaluations and benchmarking of quality of care. Since its implementation in the United States, the MDS has served as a useful tool for clinical assessment, quality monitoring, and medical reimbursement.

Yet, this U.S.-based MDS only focuses on one small segment of the aging population, i.e. older persons in institutionalized settings. Many older persons have a strong desire to maintain their autonomy and live independently even when faced with chronic illness. Moreover, in rapidly aging populations, long-term institutionalized care for older persons with mild to moderate morbidity and disability is financially unsustainable for many health care systems. These social phenomena provoke the question how to create a flexible, though standardized, tool which can assess physical and mental wellbeing and quality of life in older persons residing in both institutionalized and non-institutionalized settings. This issue is particularly pertinent in the Netherlands, where health policy is shifting towards greater reliance on informal care networks and home-based professional care as well as a more proactive, integrated approach to medicine. Such policy changes necessitate the adoption

of a standardized data collection tool measuring core outcomes in geriatric health and wellbeing to facilitate evaluation and implementation studies.

Thus, in 2008, the Dutch Ministry of Health, Welfare and Sport (VWS) established the National Care for the Elderly Programme [Nationaal Programma Ouderenzorg, NPO (ZonMw)]. This program funded over 60 research and implementation projects aimed at improving quality of life and self-reliance among older persons as well as the development of The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS). As part of TOPICS-MDS initiative, a uniform dataset measuring key outcomes in health, wellbeing, and health services utilization was created and administered to all older persons and caregivers participating in these studies. Data were centrally pooled into a public access data repository allowing researchers to explore health issues in a broader framework which may have not been possible within individual projects. Given its unique structure, TOPICS-MDS was envisioned as a catalyst for data sharing, a source for scientific and healthy policy innovation, and a tool to facilitate integrated care.

1.2. Aims of thesis

Validation refers to the accuracy of an instrument, in other words, that the instrument truly measures what it purports to measure.[4] Although instruments may be globally referred to as 'valid', there are in fact several distinct subtypes of validity. For instance, content validity refers to whether the items of an instrument comprehensively capture the underlying construct, criterion validity (sometimes referred to as convergent validity) assesses the degree of coherence between a measure and a reference standard, and construct validity evaluates whether the performance of a measure adheres to *a priori* hypotheses.[5] Validation is a critical preliminary step in research to ensure that subsequent analyses and interpretation are based on sound measurements.

Although TOPICS-MDS is comprised of several different survey instruments previously validated for use in older populations, there remains a fundamental need to re-evaluate measurement properties in the context of this data sharing initiative. Whereas previous validation work gives credence to the robustness of an instrument, the instrument itself is never truly 'valid'.[4] Rather, validity is a characteristic of an instrument for a certain population[4], underscoring the need to validate data collected as part of TOPICS-MDS.

Thus, the primary aim of this thesis was to validate the three main survey instruments included in TOPICS-MDS, a modified version of the Katz Index of Independence in Activities of Daily Living, the EQ-5D (a health-related quality

of life instrument), and the CarerQol (a care-related quality of life instrument) in the pooled NPO study population. As a secondary aim, this thesis validated two variables derived from TOPICS-MDS, a frailty index and a composite endpoint for quality of life. To complement these aims, this thesis further provides an overview of the development of TOPICS-MDS and discusses how the data can be utilized to facilitate scientific innovation and public health initiatives for older persons and their caregivers.

1.3. Thesis outline

This thesis is broadly divided into four sections: a background describing TOPICS-MDS, the evaluation of instruments included within the survey, the validation of variables derived from TOPICS-MDS, and a general discussion. Specifically:

Chapter 2 is the background chapter describing the origin and development of TOPICS-MDS and provides descriptive statistics for the database.

Chapter 3 explores the psychometric properties of the three primary survey instruments included in TOPICS-MDS (a 15-item Katz Index of Independence in Activities of Daily Living, the EQ-5D, and the CarerQol). Sections 3.1 and 3.2 examines how residential setting and survey mode impact reporting of Activities of Daily Living and Instrumental Activities of Daily Living. Section 3.3 investigates the measurement properties of the EQ-5D across four major geriatric conditions: hearing issues, urinary incontinence, joint damage, and dizziness with falls. Section 3.4 examines reporting differences across sampling frames (i.e. primary care, general population, or hospital-based sampling) and survey mode (i.e. written questionnaire and interview) for the CarerQol.

Chapter 4 describes the validation of variables derived from TOPICS-MDS. Section 4.1 validates a frailty index derived from the MDS against a frailty index based on independent, clinical assessment. Section 4.2 examines the advantage of including frailty in addition to multimorbidity and Activities of Daily Living when modelling health-related quality of life and care costs. Section 4.3 evaluates the construct validity of a preference-weighted composite endpoint for health-related quality of life developed using TOPICS-MDS.

This thesis concludes with **Chapter 5**, which contextualizes the findings of this thesis in current trends in geriatric research. Section 5.1 briefly summarizes the overall findings. Section 5.2 discusses the value of minimum datasets, standardized core outcomes, and the validation work performed as part of this thesis. Sections 5.3 and 5.4 describe the limitations and future directions in research; concluding remarks are given in Section 5.5.

References

1. Hawes C, Morris JN, Phillips CD, Fries BE, Murphy K, Mor V. Development of the nursing home Resident Assessment Instrument in the USA. *Age Ageing*. Sep 1997;26 Suppl 2:19-25.
2. Cieza A, Oberhauser C, Bickenbach J, Chatterji S, Stucki G. Towards a minimal generic set of domains of functioning and health. *BMC Public Health*. 2014;14:218.
3. Centers for Medicare & Medicaid Services. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MDSPubQIandResRep/index.html. Accessed December 10, 2015.
4. Streiner DL, Norman GR. *Health Measurement Scales: A pratical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008:167-210; 247-276.
5. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. Jan 2007;60(1):34-42.

2
Background

2.1

Less is More: Developing TOPICS-MDS

Published as:

Lutomski JE, Baars MA, Schalk BW, Boter H, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis RJ; TOPICS-MDS Consortium. The development of The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PloS One*. 2013 Dec 4;8(12):e81673. doi: 10.1371/journal.pone.0081673. eCollection 201.

Abstract

Introduction: In 2008, the Ministry of Health, Welfare and Sport commissioned the National Care for the Elderly Programme. While numerous research projects in older persons' health care were to be conducted under this national agenda, the Programme further advocated the development of The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS) which would be integrated into all funded research protocols. In this context, we describe TOPICS data sharing initiative (www.topics-mds.eu).

Materials and methods: A working group drafted TOPICS-MDS prototype, which was subsequently approved by a multidisciplinary panel. Using instruments validated for older populations, information was collected on demographics, morbidity, quality of life, functional limitations, mental health, social functioning and health service utilisation. For informal caregivers, information was collected on demographics, hours of informal care and quality of life (including subjective care-related burden).

Results: Between 2010 and 2013, a total of 41 research projects contributed data to TOPICS-MDS, resulting in preliminary data available for 32,310 older persons and 3,940 informal caregivers. The majority of studies sampled were from primary care settings and inclusion criteria differed across studies.

Discussion: TOPICS-MDS is a public data repository which contains essential data to better understand health challenges experienced by older persons and informal caregivers. Such findings are relevant for countries where increasing health-related expenditure has necessitated the evaluation of contemporary health care delivery. Although open sharing of data can be difficult to achieve in practice, proactively addressing issues of data protection, conflicting data analysis requests and funding limitations during TOPICS-MDS developmental phase has fostered a data sharing culture. To date, TOPICS-MDS has been successfully incorporated into 41 research projects, thus supporting the feasibility of constructing a large (>30,000 observations), standardised dataset pooled from various study protocols with different sampling frameworks. This unique implementation strategy improves efficiency and facilitates individual-level data meta-analysis.

Introduction

Demographic shifts towards an older population have given rise to new health care challenges across high-income countries. Despite general improvements in self-perceived health over time, health expectancy metrics have revealed increases in life years with chronic co-morbidity and mild functional impairment.[1] As health profiles of populations change, contemporary health care systems must be re-evaluated to ensure the best provision of care to older persons with more complex needs.

Such is the case for the Netherlands, where an estimated 10% of the population will be 85 years or older by the year 2050.[2] With the aim of developing a more proactive, integrated health care system to accommodate the growing number of older patients, in 2008, the Dutch Ministry of Health, Welfare and Sport commissioned the National Care for the Elderly Programme. Under this Programme, a network of local health care providers, consumer advocates and research centres was established with the guiding principles of improving care, quality of life and self-reliance among older persons. To achieve these goals, the Programme promoted research in physical, mental and social health and well-being.

While numerous research projects were to be conducted under this national agenda, the Programme further advocated the development of The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS) which would be integrated into all funded research protocols. This framework would not only have the intrinsic advantage of gathering uniform information on a large sample of older persons and caregivers at minimal cost but also promote data sharing between institutions. The Programme envisioned individual patient data could then be pooled to facilitate meta-analysis as well as serve as a public repository for external users.

Internationally, policymakers, geriatricians and other health professionals have long recognized the utility of incorporating minimal data collection as part of routine management in care facilities[3,4] and hospitals[5] as a well as a mechanism to achieve standardized outcome measurements in research.[6,7] In this context, TOPICS-MDS was developed to serve as a complementary instrument which would not only collect information on older persons but also informal caregivers and health services utilization. TOPICS-MDS therefore has a broader scope than previous minimal datasets on older persons' health and contains data relevant for many disciplines, including gerontology, public health and health economics.

Given that TOPICS-MDS was created as a large-scale data sharing initiative[8], the aim of this first paper on the database was two-fold: (1) to describe the development of TOPICS-MDS and feasibility of data collection; and (2) to discuss how frequently met challenges in building a public data repository were overcome.

Materials and methods

Project management and governance

TOPICS-MDS project was carried out as a collaborative effort between the eight medical research centers in the Netherlands, with Radboud University Medical Center serving as the central institution. A Project Group was established to advise on the development and maintenance of TOPICS-MDS and comprised of twelve members, a single representative from each medical center and four additional working group members with expertise in database management and epidemiology. To ensure the commitment of all involved parties, TOPICS-MDS project was overseen by a nationally representative Steering Committee comprised of eight stakeholders from different geographical regions within the national network.

Development of the minimal dataset instrument

Since TOPICS-MDS instrument would be incorporated into a range of research projects, it was therefore critical that the instrument was finalized prior to the commencement of these projects. Thus, the first priority of the National Care for the Elderly Programme was to develop a concise, standardized instrument which would collect essential information on the health status of the older persons and informal caregivers. Using validated instruments for use in older populations, a small working group was nominated to draft a prototype for TOPICS-MDS instrument. The working group outlined key domains and data points for the initial prototype. Several revisions of TOPICS-MDS instrument were undertaken before consensus was achieved among working group members. Upon consensus, an independent multi-disciplinary panel with expertise in gerontology, epidemiology, biostatistics and health services research was invited to evaluate the instrument's content and utility. Only minor revisions were warranted from the panel's feedback.

TOPICS-MDS was then piloted in four regions throughout the Netherlands. A descriptive analysis was conducted to identify patterns in missingness. Two main operational issues were observed; several questions were consistently misinterpreted due to either (1) linguistic construct or (2) lay-out. Thus, a plain language expert was commissioned to revise TOPICS-MDS instrument for clarity and readability, and a finalized version of the instrument was approved. The English translations of the surveys administered to older persons and informal caregivers are available at: http://www.topics-mds.eu.

Included measurements: Older persons

For older persons, information was collected on demographics, morbidity, quality

of life, functional limitations, mental health, social functioning and health service utilization for a total of 51 data points.

Demographics

The following demographic characteristics were included in TOPICS-MDS: sex; age; marital status; country of origin; primary domicile (e.g. independent residence, retirement home, nursing home); educational level and socio-economic status. In accordance with the Dutch educational system, educational level was classified into seven categories, with the lowest category representing less than primary school and the highest representing college/postgraduate education. Socio-economic status was categorized according to the Dutch Social and Cultural Planning Office Socio-Economic Status Index.[9] For this index, respondents' residential post codes were linked to geospatial data on average income, employment type and educational level to create an overall summary score, with higher numbers indicating higher socioeconomic status.

Morbidity

Respondents were asked to indicate morbidities experienced in the last twelve months from 17 pre-defined conditions (e.g. diabetes, asthma, cancer). Included conditions were based on a listing widely used in the Netherlands to record multimorbidity.[10] The presence of two or more conditions from this listing indicates multi-morbidity.

Quality of life

The EuroQol Five Dimensional scale (EQ-5D)[11] is recognized as an optimal instrument to derive preference based quality of life values, particularly when brevity is required.[12] For the purposes of this project, a modified version of the EQ-5D, the EQ-5D+C, was used.[13] Whereas the traditional EQ-5D assesses five attributes (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), the EQ-5D+C includes an additional attribute to assess cognitive function. Each attribute has three response options ('no problems', 'some problems' or 'extreme problems'), resulting in a score or '1', '2' or '3' respectively. Individual attribute scores are then concatenated into a six-digit number to describe a respondent's health state, with '111111' representing the best possible health state and '333333' the worst possible health state. Thus, this metric has the potential to describe up to 729 (3^6) unique health states. To date, there is no validated weighting formula to convert the EQ-5D+C health state to a summary index in the Dutch population.[13] However, such weightings are available for the EQ-5D[14], and EQ-5D summary scores are available in TOPICS-MDS.

Respondents were also asked to rate their current quality of life from a five-level response option ranging from 'poor' to 'excellent' and their quality of life relative to the previous year from a five-level response option ranging from 'much worse' to 'much better'. These two questions were formed using phrasing similar to self-perceived health questions from the RAND-36, which is an internationally recognized health-related quality of life survey validated for use in the Netherlands.[15,16] Self-perceived quality of life was further assessed with a modified version of Cantril's Self Anchoring Ladder[17], where respondents were asked to rate their present life on a scale between zero and ten.

Functional limitations

The extent of functional limitations was measured using a modified version of the Katz Index of Independence Basic Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL) and an additional indicator of mobility.[18] To measure functional limitations, respondents were asked if assistance is required for six basic functions (i.e. bathing; dressing; eating; toileting; use of incontinence products; getting up from a chair) and seven instrumental functions (i.e. grooming; use of telephone; travelling; grocery shopping; meal preparation, household tasks; taking medications; financial management). To measure mobility, respondents were asked if assistance was required while walking. Metrics to assess ADL and IADL, such as the Katz Index, have been administered in a variety of geriatric populations[19] and has been shown to produce reliable results irrespective of completion by a respondent or a proxy.[18,20] Responses are rated on a binary scoring system (dependent=1; independent=0) and summated, with higher scores representing greater functional limitations.

Emotional wellbeing

The Rand-36 mental health sub-scale[16] reliably measures a unidimensional concept of mental state[21], and was therefore utilized to assess psychological wellbeing in TOPICS-MDS. The sub-scale is comprised of five questions asking respondents how often in the past four weeks they have felt (1) very nervous, (2) calm and peaceful, (3) down-hearted and blue, (4) happy and (5) so down in the dumps nothing could cheer [them] up. Five-level mutually exclusive response options are available ranging from 'never' to 'always'. Positive attributes (e.g. feeling happy) are scored from zero to 100 respectively, whereas negative attributes (e.g. feeling very nervous) are reverse scored. Individual item scores are averaged and rescaled to produce a summary score between zero and 100, with higher scores indicating a more positive emotional state.

Social functioning

Social functioning was determined by a single question derived from the RAND-36.[16] Based on a five-level response option (from 'never' to 'regularly'), respondents were asked how often in the past four weeks their physical health or emotional problems had hampered their social activities.

Health services utilization

The number of hospital admissions, length of hospital stay and urgent care visits occurring in the twelve-month prior to administration of the survey were collected. Information on the frequency of home care assistance (e.g. community nurse) and temporary residence in a care home or a nursing home were also recorded.

Included measurements: Informal caregiver

For the informal caregiver, information was collected on demographics, hours of informal care and quality of life for a total of 27 data points.

Demographics

The following demographic characteristics were included in TOPICS-MDS: sex; age; socioeconomic status, the caregiver's relationship with the care recipient; whether the caregiver resided with the care recipient, and if not, the geographical distance between the caregiver and care recipient.

Hours of informal care

Respondents were asked to retrospectively indicate how many hours in the past week they assisted with household tasks, personal care, transport or financial/administrative duties. Notably, despite the potential for recall bias, retrospective reports of hours of informal care can yield valid and reliable results in cross-sectional studies if adjustments for multi-tasking are included in the analysis.[22]

Quality of Life

Similarly to older persons, self-perceived quality of life was measured using two questions adapted from the RAND-36[16] and the modified version of Cantril's Self Anchoring Ladder.[17] However, given that a high level of burden among a caregiver can negatively impact the physical and mental wellbeing of both the caregiver and care recipient[23], subjective care-related burden was measured using the validated CarerQol-7D.[24] The CarerQol-7D was modeled after the EuroQol 5-D and includes seven attributes: care-related fulfilment; relational problems with the care recipient; mental health; time management; financial security; social support; and physical health. From three response options

('no', 'some', 'a lot', scored as '1', '2' and '3' respectively), respondents can indicate the extent of each attribute in their personal situation.[24] The CarerQol-7D score is derived likewise to the EuroQol 5-D+C[13] and thus can describe up to 2,187 (3^7) levels of care-related burden.

Care-related burden was further assessed with the CarerQol-VAS[24], which uses a visual analogue scale ranging from '0' to '10' to rate a caregiver's level of happiness from 'completely unhappy' to 'completely happy'. Difficulty of care provision and the level happiness if another were to assume care responsibilities were also rated with a VAS.

Sampling framework and longitudinal data collection

In total, 52 independent research studies included TOPICS-MDS into their research protocols. The study design, sampling framework and inclusion criteria differed across research studies. Several individual project protocols included longitudinal data collection. In such cases, TOPICS-MDS instrument was administered at baseline and at least one additional follow-up was scheduled, typically 12 months after baseline.

Ethical approval

TOPICS-MDS instrument was integrated into pre-existing research protocols, and therefore ethical approval for the collection of TOPICS-MDS was sought from individual study sites. Results presented in this analysis were exempt from institutional review as data were anonymised and within the public domain.

Data collection and management

A data dictionary and a standardized protocol for data cleaning procedures were drafted and provided to all participating project managers. To preserve participant confidentiality, data were cleaned at individual research sites, stripped of any personal identifiers and entered into a standardized computerized database. All data were submitted to a centralized body (Radboud University Medical Center) for the collation of a national dataset.

Development of a public data repository

To facilitate external users, all de-identified data maintained in the public repository have been verified for accuracy and clearly labelled. A single institution (Radboud University Medical Center) was nominated to be the custodian of TOPICS-MDS and facilitate incoming data requests. To ensure equitable use, the Project Group drafted a data access policy and selected a Societal Board to review the societal merits and benefits of all data requests.

Members of the Societal Board were nominated by the National Care for the Elderly network and comprised of seven members: a consumer advocate (chair), two research scientists, two community representatives and two health policy professionals.

Following data collection, there is a six-month moratorium in the release of the data. After this period, TOPICS-MDS data access policy permits all research scientists affiliated with an academic, healthcare or other research institution worldwide to submit a request to access data. Prior to the release of data, all requests must be approved by both the Project Group and Societal Board.

Table 1 Characteristics of 41 projects included in The Older Persons and Informal Caregivers Minimum Data Set, Netherlands, 2012

	Projects (N=41)	**Participants (N=32,310)**
Sampling frame	N	N (%)
Primary care practice	20	16,537 (51.2)
General population	10	6,401(19.8)
Hospital	5	1,753 (5.4)
Retirement community	3	5,083 (15.7)
Nursing home	3	2,536 (7.9)
Inclusion criteria[a]		
Frailty[b]	6	8,832 (27.3)
Dementia	3	2,352 (7.3)
Age minimum		
45 years	1	1,479 (4.6)
50 years	1	535 (1.7)
60 years	4	1,661 (5.1)
65 years	14	8,800 (27.2)
70 years	6	1,688 (5.2)
75 years	6	10,876 (33.7)
Unspecified	9	7,271 (22.5)

[a] Inclusion criteria presented in this table are neither exhaustive nor mutually exclusive.
[b] Operational definitions for frailty differed across studies.

Results

TOPICS-MDS contains essential data to better understand health challenges experienced by older persons and informal caregivers. To highlight the utility of the TOPICS-MDS, descriptive statistics were calculated using preliminary data from 41 research projects. Results are shown for select characteristics only. All analyses were performed using SAS 9.2 (Carey, NC, USA).

Between 2010 and 2013, a total of 41 research projects contributed data to TOPICS-MDS. The majority of studies sampled were from primary care settings and inclusion criteria differed across studies (Table 1). The following analyses are based on the preliminary data for 32,310 older persons and 3,940 informal caregivers.

The majority of older persons in this cohort were women (59.0%). Relative to men, women were modestly older, more likely to be widowed and more likely to reside independently (Table 2). Multi-morbidity was common, with three-quarters of respondents reporting two or more morbidities. The most frequently cited conditions were hearing problems (45.8%), knee and hip joint damage (42.0%) and vision disorders (39.4%).

Based on the EQ-5D+C, the vast majority reported no problems with self-care; though, severe problems with anxiety and depression were relatively high among both men and women (Table 3). One-third (32.9%) of respondents reported no functional limitations in ADL or IADL. However, substantial gender disparities were observed, with women less likely to report no functional limitations than men (25.2% versus 47.5% respectively). The prevalence of incontinence products use and requiring assistance with household tasks and walking were notably higher among women (Table 4).

The average age of informal caregivers in this cohort was 63 years (SD 13); more than two-thirds (69.8%) were women. Caregivers were most likely to be a spouse/life partner or a daughter/son (in-law) of the care recipient (42.5% and 40.2% respectively); more than half (52.3%) did not live with their care recipient. Approximately 10% of caregivers reported 'some' or 'a lot' of care-related financial burden and nearly one-third (30.6%) received no social support from family, friends or acquaintances (Table 5).

Table 2 Demographic characteristics of older persons, The Older Persons and Informal Caregivers Minimum Data Set, Netherlands, 2012

	Men (N=13,237)	Women (N=19,017)
Age (mean, SD)	77 (8)	79 (8)
Marital status		
Married	70.7	35.8
Widowed	18.3	50.4
Other[a]	11.1	13.8
Dutch origin	91.5	91.1
Primary domicile		
Independent residence	27.5	50.7
Residence with family members	51.2	29.7
Retirement home	19.9	17.0
Nursing home	1.4	2.5
Educational level[b]		
Primary school or less	25.6	42.2
Practical/secondary vocational training	47.3	44.9
Some college/university degree	27.2	12.9

Note: Values are presented as percentage unless otherwise stated. Data are based on 41 research projects.

[a] Includes single, divorced and cohabiting.

[b] Collapsed from seven Dutch educational categories.

Table 3 Percentage of older persons reporting no, some or severe problems on the EQ-5D+C quality of life scale by sex, The Older Persons and Informal Caregivers Minimum Data Set, Netherlands, 2012

	No problem	Some problems	Severe problems
Mobility			
Men	50.6	47.9	1.5
Women	37.8	60.3	1.9
Self-care			
Men	83.3	12.4	4.3
Women	77.2	16.5	6.2
Usual activities			
Men	63.4	26.5	10.1
Women	53.0	36.4	10.6
Pain/discomfort			
Men	48.7	45.6	2.3
Women	33.9	55.4	10.9
Anxiety/depression			
Men	71.0	18.3	10.7
Women	62.2	25.1	12.7
Cognitive functioning			
Men	64.9	32.9	2.2
Women	67.9	30.2	2.0

Note: Percentages are based on 13,237 men and 19,017 women.

Table 4 Percentage of older persons requiring assistance for activities of daily living by sex, The Older Persons and Informal Caregivers Minimum Data Set, Netherlands, 2012

	Men (N=13,237)	Women (N=19,017)
Basic activities		
Bathing or showering	14.6	21.3
Dressing	10.5	13.6
Eating	3.3	3.1
Toileting	5.2	5.6
Use of incontinence products	11.9	39.8
Getting up from a chair	7.2	10.1
Instrumental activities		
Grooming	6.0	6.1
Use of telephone	8.7	6.4
Travelling	23.8	40.8
Grocery shopping	20.2	36.5
Preparing a meal	26.5	23.3
Household tasks	39.6	62.5
Taking medication	13.1	12.5
Financial management	16.1	20.8
Mobility		
Walking	18.9	31.9

Table 5 Percentage of caregivers reporting problems on the CarerQol-7D quality of life scale (n=3,940), The Older Persons and Informal Caregivers Minimum Data Set, Netherlands, 2012

	No	Some	A lot
Satisfaction performing care tasks	5.9	41.0	53.1
Relational problems with care recipient	52.6	34.2	13.3
Issues with personal mental health	49.9	37.1	12.9
Issues with personal physical health	44.3	40.5	15.2
Problems combining daily activities and care tasks	49.6	38.3	12.1
Financial problems	90.7	7.3	2.0
Social support in care tasks	30.6	39.0	30.5

Discussion

TOPICS-MDS has been successfully incorporated into numerous research projects, thus supporting the feasibility of constructing a large (>30,000 observations), standardized dataset pooled from various study protocols. These initial positive findings are encouraging to future researchers who may wish to administer TOPICS-MDS instrument within their own research protocols to further this initiative. The unique implementation strategy of TOPICS-MDS has several inherent strengths. First, integrating a standardized data collection tool into pre-existing research protocols is a highly efficient and cost-effective method to generate data on a large number of respondents. Moreover, by collecting uniform individual-level data, we counter traditional obstacles that impede meta-analysis, such as select reporting of aggregate data or differences in exposure/outcome operational definitions.[25,26] Lastly, with appropriate statistical considerations, the pooled data from TOPICS-MDS may be able to have broader generalizability than individual research studies.[27]

Preliminary analyses of TOPICS-MDS revealed that a considerable proportion of recruited older persons experienced some form of disutility, whether related to morbidity, ADL or quality of life, thus alluding to the underlying extent of frailty. Defined as an increased vulnerability to adverse health outcomes following a stressor event[28], frailty can serve as a stronger indicator for geriatric intervention than chronological age.[29] For this reason, identifying frailty on a patient-level can result in more effective case management, and on a population-level, can lead to improved distribution of health services. Although there are several valid methods to measure frailty[28], the calculation of a frailty index based on the accumulation of deficits in health[30] (i.e. symptoms, morbidities and/or functional limitations) can be easily applied in large-scale population studies.[31-33] Thus, given the range of deficits captured within TOPICS-MDS, a frailty index can be derived to provide another important indicator of health in the database.

TOPICS-MDS is not without limitations. Arguably, alternative metrics to those included may have permitted a more detailed investigation of outcomes of interest. However, TOPICS-MDS data collection instrument was designed to achieve a critical balance between content and succinctness. Moreover, although the data collection instrument was comprised of well-established health scales, given the different sampling frameworks of individual research projects, further methodological investigations are necessary to assess if reliability, validity and generalizability are upheld in the overall sample population. Specifically, future studies examining cluster effects, heterogeneity and patterns in missingness are warranted to maximize the utility and interpretation of the data.

Still, TOPICS-MDS should not only be seen as an endeavor to create a minimal dataset in older persons' health and wellbeing but also as a large-scale data sharing initiative, which in itself is an important scientific output. Data sharing has the potential to provoke positive changes in public health strategies, improve project cost-effectiveness and enhance scientific integrity.[34] These advantages have become increasingly recognized throughout scientific communities, consequently prompting 17 major funders of public health research to draft a joint statement supporting public data repositories.[34,35] Nonetheless, while advances have been made in biomedical spheres, data sharing remains largely elusive in public health research.[35] Commonly cited barriers include data protection legislation, potential overlap in analyses and funding limitations.[36] Other underlying issues, such as self-perceived proprietorship over databases and the competitive demand to increase published output, also contribute to a research culture which is not conducive with data sharing.[35,36] Thus, despite the potential for increased citation rates[37] and journal policies advocating open access, the release of data is not always achieved in practice.[38,39]

In light of these complications, TOPICS-MDS Project Group and Steering Committee sought to proactively address potential obstacles in order to encourage a culture of data sharing from the initial phases of the project. Firstly, to comply with data protection legislation, external users will only be permitted to access a fully anonymised database. To circumvent issues related to publication rights, a brief moratorium in the release of data is implemented to afford research consortium members the opportunity to publish without conflicting data requests. Following this period, all Project Group members acceded that they would have to submit a data request to perform any additional analyses not initiated during the moratorium. To further protect the interests of external users, TOPICS-MDS Societal Board was established as a safeguard against preferential release of data. Lastly, like many public health research projects, TOPICS-MDS received fixed funding. To promote the continuance of the project, funding calls are being actively sought by the Project Group and it is envisioned that TOPICS-MDS will be incorporated into future studies on older persons' health. Opportunities to link data with permanently funded institutions are also being explored.

Nonetheless, while these aforementioned measures are fundamental for data sharing, we believe that the strong commitment of all involved stakeholders underlies the success of this public data repository to date. Based on our experience, building TOPICS-MDS without collaborating with the researchers who collected the data would have been ineffective. Rather, we found keeping researchers engaged through regular updates and assistance with the data

submission process were instrumental to the sustainability of the data sharing initiative.

TOPICS-MDS will be open for external requests in the last quarter of 2013; full details on how to submit a request will be made available through TOPICS-MDS website at: http://www.topics-mds.eu. Additional background information, the TOPICS-MDS data dictionary and relevant syntaxes can also be accessed on the website. Documents are available in English and Dutch.

In conclusion, TOPICS-MDS represents a strong example of a public data repository with wide reaching potential. Understanding the health challenges experienced by older persons and informal caregivers can help inform the re-configuration of contemporary care models to achieve a more integrated and proactive health services system. Although based in the Netherlands, such findings are timely and relevant for many industrialized countries where increasing health-related expenditure has necessitated the evaluation of contemporary health care delivery.

References

1. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. *Lancet*. Oct 3 2009;374(9696):1196-1208.
2. van der Bie R, Latten J. *Baby Boomers in the Netherlands*. The Hague, Netherlands: Statistics Netherlands; 2012.
3. Morris JN, Hawes C, Fries BE, et al. Designing the national resident assessment instrument for nursing homes. *Gerontologist*. Jun 1990;30(3):293-307.
4. Sgadari A, Morris JN, Fries BE, et al. Efforts to establish the reliability of the Resident Assessment Instrument. *Age Ageing*. Sep 1997;26 Suppl 2:27-30.
5. Borchelt M, Vogel W, Steinhagen-Thiessen E. [The Geriatric Minimum Data Set (Gemidas) of the Federal Association of Clinical Geriatric Facilities e. V. as an instrument for quality assurance in inpatient geriatrics]. *Z Gerontol Geriatr*. Feb 1999;32(1):11-23.
6. Abellan Van Kan G, Sinclair A, Andrieu S, et al. The geriatric minimum data set for clinical trials (GMDS). *J Nutr Health Aging*. Mar 2008;12(3):197-200.
7. Ellul J, Watkins C, Barer D. Estimating total Barthel scores from just three items: the European Stroke Database 'minimum dataset' for assessing functional status at discharge from hospital. *Age Ageing*. Mar 1998;27(2):115-122.
8. Melis RJ, Vehof H, Baars L, Rietveld MC, Olde Rikkert MG. Sharing of research data. *Lancet*. Dec 10 2011;378(9808):1995.
9. Sociaal en Cultureel Planbureau. Statusscores (*available in Dutch only*). Available at: http://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores. Accessibility verified May 13, 2013.
10. Lokale en nationale monitor gezondheid (*in Dutch only*). Available at: https://www.monitorgezondheid.nl/gezondheidindicatoren.aspx. Accessibility verified March 13, 2013.
11. Brooks R. EuroQol: the current state of play. *Health Policy*. Jul 1996;37(1):53-72.
12. Haywood KL, Garratt AM, Fitzpatrick R. Quality of life in older people: a structured review of generic self-assessed health instruments. *Qual Life Res*. Sep 2005;14(7):1651-1668.
13. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol*. Apr 1999;52(4):293-301.
14. Lamers LM, Stalmeier PFM, McDonnell J, Krabbe PFM, van Busschbach JJ. Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief. *Ned Tijdschr Geneeskd*. 2005;149:1574-1578.
15. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med*. Jul 2001;33(5):350-357.
16. van der Zee K, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding. (*available in Dutch only*). Available at: http://www.rug.nl/research/share/research/tools/assessment-tools/rand36. Accessibility verified February 26, 2013.
17. Cantril H. *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press; 1965.
18. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc*. Jun 1992;40(6):585-588.
19. Palmer M, Harley D. Models and measurement in disability: an international review. *Health Policy Plan*. Aug 2012;27(5):357-364.
20. Buurman BM, van Munster BC, Korevaar JC, de Haan RJ, de Rooij SE. Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. *J Clin Epidemiol*. Jun 2011;64(6):619-627.
21. Moorer P, Suurmeije Th P, Foets M, Molenaar IW. Psychometric properties of the RAND-36 among three chronic diseases (multiple sclerosis, rheumatic diseases and COPD) in The Netherlands. *Qual Life Res*. 2001;10(7):637-645.

22. van den Berg B, Spauwen P. Measurement of informal care: an empirical study into the valid measurement of time spent on informal caregiving. *Health Econ*. May 2006;15(5):447-460.
23. Deeken JF, Taylor KL, Mangan P, Yabroff KR, Ingham JM. Care for the caregivers: a review of self-report instruments developed to measure the burden, needs, and quality of life of informal caregivers. *J Pain Symptom Manage*. Oct 2003;26(4):922-953.
24. Brouwer WB, van Exel NJ, van Gorp B, Redekop WK. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Qual Life Res*. Aug 2006;15(6):1005-1021.
25. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.
26. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. Apr 19 2000;283(15):2008-2012.
27. Stukel TA, Demidenko E, Dykes J, Karagas MR. Two-stage methods for the analysis of pooled data. *Stat Med*. Jul 30 2001;20(14):2115-2130.
28. Clegg A, Young J, Iliffe S, Olde Rikkert M, Rockwood K. Frailty in elderly people. *Lancet*. Mar 2 2013;381(9868):752-762.
29. Schuurmans H, Steverink N, Lindenberg S, Frieswijk N, Slaets JP. Old or frail: what tells us more? *J Gerontol A Biol Sci Med Sci*. Sep 2004;59(9):M962-965.
30. Mitnitski AB, Mogilner AJ, MacKnight C, Rockwood K. The mortality rate as a function of accumulated deficits in a frailty index. *Mech Ageing Dev*. Sep 2002;123(11):1457-1460.
31. Romero-Ortuno R, Kenny RA. The frailty index in Europeans: association with age and mortality. *Age Ageing*. Sep 2012;41(5):684-689.
32. Larsen KD, Rubinfeld IS. Changing risk of perioperative myocardial infarction. *Perm J*. Fall 2012;16(4):4-9.
33. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. Aug 30 2005;173(5):489-495.
34. Wellcome Trust. Public health research data forum. Available at: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030689.htm. Accessibility verified March 5, 2013.
35. Walport M, Brest P. Sharing research data to improve public health. *Lancet*. Feb 12 2011; 377(9765): 537-539.
36. Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *Lancet*. May 14 2011;377(9778):1633-1635.
37. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLoS One*. 2007;2(3):e308.
38. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One*. 2009;4(9):e7078.
39. Eysenbach G, Sa ER. Code of conduct is needed for publishing raw data. *BMJ*. Jul 21 2001; 323(7305):166.

3

Evaluation of TOPICS-MDS Survey Instruments

3.1

The influence of residential setting on ADL and IADL reporting

Accepted for publication as:

Lutomski JE, Krabbe PFM, den Elzen WPJ, Olde-Rikkert MGM, Steyerberg EW, Muntinga ME, Bleijenberg N, Kempen GIJM, Melis RJF on behalf of TOPICS Consortium. Comparative analyses of ADL/IADL summary scores from different residential settings is inappropriate. *Journal of Clinical Epidemiology*. (In press)

Abstract

Objective: To internally validate a 15-item dichotomous Activities of Daily Living and Instrumental Activities of Daily Living (ADL/IADL) Index.
Methods: Data were extracted from The Older Persons and Informal Caregivers Survey Minimal DataSet (TOPICS-MDS). Using Rasch modelling, six aspects of the ADL/IADL scale were assessed: (1) overall fit; (2) internal consistency; (3) individual item and person fit; (4) local dependency; (5) targeting; and (6) differential item functioning (RUMM 2030). All analyses were stratified by living situation [community-dwelling (n=21,926) or residential care facility (n=2,458)].
Results: In both settings, 'eating' was the easiest activity on the scale and 'performing household tasks' was the most difficult activity. However, based on the location on the logit scale, the level of difficulty for certain items varied between residential settings, suggesting summary scores are not equivalent between these settings. Differential item functioning by gender and age group was observed for several items, indicating potential measurement bias in the scale.
Discussion: Unless adjustments are undertaken, ADL/IADL summary scores retrieved from older persons residing in the community or residential care facilities should not be directly compared. This 15-item scale is poorly targeted for a community-dwelling older population, underscoring the need for items with improved discriminative ability.

Introduction

The concept of validity is a cornerstone in clinical and public health research. When measuring subjective health outcomes, researchers actively promote the use of validated instruments. Importantly, however, such instruments are not inherently valid, but rather validated in the context of specific research settings.[1] Survey methodologists have therefore supported the re-evaluation of instrument validity in different settings.[1]

There are two major measurement paradigms to assess validity: Classical Test Theory (CTT) and Item Response Theory (IRT). CTT approaches are often applied in epidemiological and clinical validity research. The central tenet of CTT is that any observed score is a reflection of a "true" score and a degree of random error.[2] However, recognized criticisms of this paradigm include the underlying assumptions that there is a constant error term (i.e. measurement error is independent of the true score), items on a scale are equally weighted, and results are sample-dependent.[2]

In contrast, IRT is a probabilistic model of how individuals respond to items[2] and is based on the assumption that an outcome is a function of the item's characteristics and the individual completing the survey item.[1] Although historically rooted in psychometrics, IRT approaches have been increasingly applied in public health research[3] and have been regarded as a useful tool to assess validity.[4,5] Unlike CTT, IRT does not assume that all scale items are equally weighted[2], a highly useful aspect in validation work. IRT approaches can further investigate whether a scale covers a sufficient range of items (supporting the measurement of a larger theoretical construct) and if meaningful variation in response patterns are present across demographic groups (indicating potential measurement bias).[6] Within IRT, the adherence to the Rasch model represents a robust validity test since this specific IRT methods assesses person and health-state parameters separately.[7]

Given these aforementioned properties, IRT confers clear advantages when assessing the validity of Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) scales. Briefly, ADL and IADL represent activities associated with independent living, particularly among older persons. The former concept refers to more basic activities, such as eating and toileting, whereas the latter refers to more complex activities, such as shopping and managing finances. Since diminished ability to complete such activities may result in poorer quality of life[8] and increased risk for depression[9], ADL/IADL scores are important health markers.[10] Numerous scales have been developed to measure ADL/IADL[11,12], with one review of the literature identifying more than one hundred variations.[13] When exploring the validity of a scale, it is important

to acknowledge that ADL/IADL scales possess a hierarchy of difficulty; in other words, some items are inherently more difficult to perform than others[14,15] For this reason, IRT approaches can elucidate additional information which is not possible using CTT approaches.

The primary aim of our study was to internally validate a 15-item dichotomous ADL/IADL index[16] used in The Older Persons and Informal Caregivers Survey Minimal DataSet (TOPICS-MDS), a large data sharing initiative containing prospectively collected data in the Netherlands. IRT approaches, specifically Rasch analyses, have been previously applied to different ADL/IADL scales.[17,18] We therefore complement this research by applying a Rasch analysis to this specific 15-item dichotomous ADL/IADL scale and further compare the validity of this scale across two residential settings (community-dwelling and residential care facility) to inform future users on the appropriateness of comparing and/or combining summary scores from these settings. This aim is particularly relevant in the context of TOPICS-MDS because this initiative was developed, in part, to generate comparative outcomes between contributing research projects.

Methods

Data source

Data were derived from TOPICS-MDS, which is a public access data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere[19]; the original survey used for data collection is available on TOPICS-MDS website (www.topics-mds.eu). Briefly, the Dutch National Care for the Elderly Programme (http://www.nationaalprogrammaouderenzorg.nl) was established in 2008 to promote proactive, integrated health care for older persons with complex care needs. As part of this national agenda, TOPICS-MDS was developed to prospectively collect uniform information from all research projects funded under the Programme. Included survey items were based on the recommendations of an expert panel who was tasked with identifying key outcomes in older persons' health relevant in a range of settings.[19]

TOPICS-MDS consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardized protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to

TOPICS-MDS, the database is dynamic in nature and may be updated with new observations. Our present analysis uses data from 41 research projects with complete ADL/IADL data available on 24,648 older persons. TOPICS-MDS is a fully anonymized dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Centre Ethical Committee review reference number: CMO: 2012/120).

Measures

Activities of daily living

A modified version of the Katz Index of Independence of Activities in Daily Living[16] was reviewed by a multidisciplinary expert panel convened during the development phase of TOPICS-MDS.[19] This scale was selected since it provided a more global measure of limitations in activities and allowed for ADL and IADL to be analyzed separately as well as combined. Participants were asked if they currently required assistance performing 15 different activities. ADL included bathing, dressing, eating, toileting, use of incontinence products, and getting up from a chair (transferring). IADL included grooming, use of telephone, travelling, shopping, meal preparation, household tasks, taking medications, financial management, and walking. All responses were recorded dichotomously (yes, assistance required/no assistance required). Items were summated, resulting in a scale ranging from 0 to 15 limitations.

Demographics

We decided *a priori* to stratify all analyses by living situation, defined as either community-dwelling or residing in a residential care facility. In the Netherlands, a residential care facility provides housing and care for older persons who encounter difficulties living independently but do not require the high level of care provided in a nursing home setting. Residents are encouraged to perform ADL and IADL; however, assistance with activities is available. Age was divided into three groups (65 – 74; 75 – 84; 85+).

Analysis

From the class of IRT models, we applied the most basic model, the Rasch model. This IRT model applies a 1-parameter logistic model to assess an item's discriminative ability. The model assumes that the probability of item endorsement is a logistic function of the relative distance between the item's difficulty ("item location") and the participant's ability ("person location").[4] Thus, in the context of the present study, the probability that a participant reports an ADL/IADL limitation is interpreted as a logistic function of the difficulty of the activity relative to the person's overall ability.

Previous studies have demonstrated that combined ADL/IADL represent a unidimensional latent construct of disability[15,20-22], a requirement for Rasch analysis. To ensure this requirement was upheld, we first re-assessed the dimensionality of the scale in both residential settings. In line with previous research[4,23,24], we then evaluated six aspects to assess the properties of the ADL/IADL scale: (1) overall fit to the Rasch model; (2) internal consistency; (3) individual item and person fit; (4) local dependency; (5) targeting; and (6) differential information functioning. Data analyses were performed using R (2013: Vienna, Austria) and Rasch Unidimensional Measurement Model 2030 (RUMM2030; Rumm Laboratory, Perth, Western Australia).

Endorsement patterns

Based on frequencies, we first examined the endorsement patterns for each ADL/IADL item.

Unidimensionality

Using the R package 'lavaan'[25], we conducted a confirmatory factor analysis to evaluate the unidimensionality of the scale. Goodness-of-fit was examined using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI). For the confirmatory factor analysis, a RMSEA <0.08[26], a CFI >0.90[27] and a TLI >0.90[28] were considered the minimum thresholds of acceptability for model fit.

Overall fit to the Rasch model

We examined the standardized mean item and person fit residuals and the χ^2 coefficient. The residuals indicate the level of agreement between the observed and expected responses given the ability of the individual and the difficulty of the item; perfect fit would be represented with a mean residual of zero with a standard deviation falling between negative one and one.[4] A standard deviation >1.5 indicates poor model fit.[4]

The χ^2 value represents item-trait interaction, whether items follow the same hierarchical order of difficulty across class intervals (i.e. groups) on the measurement continuum. We tested the significance of the χ^2 value using a Bonferroni-adjusted α to allow for multiple comparisons; a non-significant χ^2 value indicated good fit. However, large sample sizes (>1,000) inflate the χ^2 value, and increase the likelihood of identifying misfit.[29] We therefore calculated the RMSEA as a complementary fit indicator. The RMSEA is less influenced by sample size[29] and was derived using the following formula:

$$\text{RMSEA} = \sqrt{[((\chi^2/df) - 1)/(N - 1)]}$$

whereby χ^2 is the chi-square value, df is the associated degrees of freedom and N is the sample size. We considered an RMSEA value <0.2 to indicate adequate model fit.[29] Notably, the threshold for the RMSEA differs between assessing unidimensionality in a confirmatory factory analysis and assessing overall fit for a Rasch model.

Internal consistency

Internal consistency was evaluated using the Person Separation Index; this index indicates the extent to which the ADL/IADL scales discriminate between persons of different ability. When calculating the Person Separation Index, extreme scores (i.e. persons who reported having all ADL/IADL limitations or no ADL/IADL limitations) produce less precise estimates. We therefore calculated this index with and without extreme scores. Similar to Cronbach's α coefficient, a Person Separation Index score >0.70 is recommended for scales interpreted at the population level.[4] Scales with index scores lower than this threshold may require additional items or response options to better discriminate between persons with different ADL/IADL levels.

Individual item location and fit

We evaluated individual item location and fit residuals. Item locations are presented in logits, i.e. log-odds units, a unit of measurement representing the difference between individual ability and item difficulty.[30] Extreme positive fit residuals (>2.5) suggest misfit whereas extreme negative fit residuals (<-2.5) suggest item redundancy.[4]

Local dependency

Rasch analyses require that items are locally independent of each other. This means that there should not be any correlation between two items after the effect of the underlying trait is conditioned out, i.e. the correlation should be zero[16] We classified items as locally dependent if the correlation coefficient between their residuals was >0.3.[4] If locally dependent items were observed, the items were examined to evaluate their impact on the reliability of the scale.

Targeting

Targeting is defined as the extent to which the range of the measure corresponds with the range of the measure in the study sample.[5] If a scale is poorly targeted, i.e. too easy or too difficult, persons may disproportionately report having no ADL/IADL limitations (floor effect) or all ADL/IADL limitations (ceiling effect). To assess targeting, we examined the person-item location distribution map, which visually depicts person locations against item locations. A positive mean

location score indicates that the study population demonstrates a higher ADL/IADL ability than the average of the scale whereas a negative mean location score indicates the contrary.[4,31] A mean location score near zero indicates a well-targeted scale.

Differential item functioning

To assess differential item functioning (DIF) across gender and age group, we first performed an ANOVA of the standardized response residuals of each item by class interval. A significant main effect for gender or age group indicates uniform DIF, i.e. a consistent systematic difference between groups across the measurement continuum. A significant interaction between gender/age group and class interval indicates non-uniform DIF, i.e. an inconsistent difference between groups. To confirm DIF, we then visually inspected the observed item characteristic curves by gender and age group relative to the theoretical item characteristic curve.

Results

Endorsement patterns

Based on frequencies, older persons in both settings were most likely to report having difficulties with 'household tasks' and least likely to report have difficulties with 'eating' (Table 1).

Unidimensionality

Unidimensionality was upheld in the community-dwelling and residential care facility settings (Table 2), thus meeting the prerequisite for performing a Rasch model.

Overall fit to the Rasch model

The level of adherence to the Rasch model varied by setting, with the community-dwelling setting exhibiting poorer fit (Table 3). The standardized mean fit item residual was notably large for the community-dwelling setting; the item residual standard deviations exceeded an acceptable threshold (i.e. >1.5) in the community-dwelling and residential care facility settings. Based on the χ^2 test, a statistically significant item-trait interaction was present in both settings. However, when cross-referenced with the RMSEA to counter the influence of large sample size, the fit was reasonable (i.e. RMSEA <0.2). Mean standardized person fit residuals and corresponding standard deviations fell within an acceptable range (Table 3).

Table 1 Characteristics of study population, The Older Persons and Informal Caregivers Survey Minimum Dataset, 2013

	Community-dwelling N=21,926 %	Care facility N=2,458 %
Female	58.1	73.7
Age group (years)		
65 – 74	32.1	4.7
75 – 84	52.2	37.0
>85	15.7	58.3
Required assistance with ADL/IADL items		
Household tasks	48.0	89.9
Travelling	27.7	69.9
Meal preparation	16.8	67.0
Shopping	24.8	62.4
Walking	22.5	58.4
Bathing*	11.6	58.4
Use of incontinence products*	25.1	52.8
Financial management	14.5	42.6
Dressing*	8.0	37.5
Taking medications	6.8	31.5
Getting up from a chair*	6.7	18.0
Toileting*	2.7	13.7
Grooming	2.2	12.7
Use of telephone	3.9	11.8
Eating*	1.3	4.0

Abbreviations: ADL, activities of daily living; IADL, instrumental activities of daily living
Note: * indicates ADL item.

Table 2 Confirmatory factor analysis model fit by residence setting

	Community-dwelling N=21,926	Residential care facility N=2,458
χ^2	3,841	1,087
df	90	90
χ^2 *p*-value	<0.001	<0.001
RMSEA (95% CI)	0.044 (0.042-0.050)	0.067 (0.064-0.071)
CFI	0.989	0.969
TLI	0.987	0.963

Abbreviations: df, degrees of freedom; RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker-Lewis Index
Note: χ^2 value refers to the minimum function test statistic. An RMSEA <0.08, a CFI >0.90, and a TLI >0.90 were considered the minimum thresholds of acceptability.

Internal consistency

The Person Separation Index was sub-optimal for the ADL/ADL scale in the community-dwelling setting (0.54 with extreme scores; 0.61 without extreme scores). This finding implies that the scale does not discriminate well between persons of different abilities in this setting. However, the Person Separation Index was acceptable (i.e. >0.70) in the residential care facility setting irrespective of the inclusion or exclusion of extreme scores (Table 3).

Individual item location and fit

Tables 4 show the item locations for the combined ADL/IADL scale. In both settings, 'eating' was the easiest activity on the scale and 'performing household tasks' was the most difficult activity. However, based on the location on the logit scale, the level of difficulty for certain items varied between settings. For example, 'bathing' and 'meal preparation' represented easier items in the community-dwelling setting (0.12 and 0.80 logits respectively) than in the residential care facility setting (1.22 and 1.62 logits respectively).

In the community-dwelling setting, twelve of the 15 items had a variance that did not fit with the Rasch model (i.e. fit residuals outside ±2.5 range). Most suggested item redundancy (i.e. negative fit residuals). Two items, 'use of incontinence products' and 'managing finances', displayed misfit (i.e. positive fit residuals). In the residential care facility setting, eight items exhibited poor fit; five items suggested item redundancy whereas three items ('managing finances', 'use of incontinence products', and 'walking') suggested misfit.

Table 3 Overall scale and person fit to the Rasch Model

Residence type	Standardized mean item fit residual	SD	χ^2	p-value	RMSEA	Mean standardized person fit residual	SD	PSI with extremes	PSI without extremes
Community-dwelling	-4.84	15.60	7309.40	<0.0001	0.06	-0.32	0.52	0.54	0.61
Residential care facility	-0.87	3.82	650.98	<0.0001	0.04	-0.28	0.54	0.81	0.77

Abbreviations: sd, standard deviation; RMSEA, root mean square error of approximation; PSI, person separation index
Note: A mean fit residual of 0 with a standard deviation between 1 and -1 suggest perfect fit. Whereas a non-significant χ^2 value typically indicates good fit, in large samples, a RMSEA value <0.2 indicates adequate model. A PSI score >0.70 is recommended for scales interpreted at the population level.

Table 4 Item fit statistics for ADL/IADL items in community-dwelling setting (N=21,926) and residential care facility setting (N=2,458), ordered by item location

Item	Location	SE	Fit residual	χ^2	p-value
Community-dwelling					
Eating	-2.97	0.06	-1.91	15.58	0.02
Grooming	-2.44	0.05	-9.07	128.95	<0.0001*
Toileting	-2.19	0.05	-10.68	197.85	<0.0001*
Telephone	-1.42	0.04	-0.90	58.08	<0.0001*
Medicine	-0.69	0.03	-5.93	64.52	<0.0001*
Getting up from chair	-0.68	0.03	-4.44	53.49	<0.0001*
Dressing	-0.50	0.03	-17.37	511.98	<0.0001*
Bathing	0.12	0.02	-22.85	726.89	<0.0001*
Managing finances	0.55	0.02	18.45	980.86	<0.0001*
Meal preparation	0.80	0.02	-9.72	255.99	<0.0001*
Walking	1.31	0.02	-5.50	183.85	<0.0001*
Use of incontinence products	1.42	0.02	36.99	2086.78	<0.0001*
Shopping	1.53	0.02	-24.24	977.82	<0.0001*
Travelling	1.79	0.02	-16.10	849.43	<0.0001*
Household tasks	3.36	0.02	0.69	217.35	<0.0001*
Residential care facility					
Eating	-3.86	0.12	-1.98	9.30	0.41
Grooming	-2.16	0.07	-3.22	36.19	<0.0001*
Toileting	-2.09	0.07	-4.27	81.76	<0.0001*
Telephone	-2.07	0.07	1.44	34.23	<0.0001*
Getting up from chair	-1.45	0.06	-0.59	21.26	0.01
Medicine	-0.46	0.05	1.93	16.13	0.06
Dressing	-0.10	0.05	-7.22	112.99	<0.0001*
Managing finances	0.22	0.05	3.75	18.54	0.03
Use of incontinence products	0.79	0.05	7.01	52.02	<0.0001*
Walking	1.13	0.05	2.80	13.93	0.12
Bathing	1.22	0.05	-5.97	114.18	<0.0001*
Shopping	1.36	0.05	-3.78	77.82	<0.0001*
Meal preparation	1.62	0.05	-0.34	20.30	0.02
Travelling	1.87	0.05	-1.47	31.45	0.0002*
Household tasks	3.97	0.09	-1.14	10.89	0.28

Abbreviation: ADL, activities of daily living; IADL, instrumental activities of daily living; SE, standard error
Note: *Bonferroni adjusted significant level, 0.000667

The majority of items demonstrated statistically significant item-trait interactions (χ^2 test), i.e. inconsistent performance across class intervals of difficulty, in the community-dwelling and residential care facility settings. These findings were likely inflated due to large sample size.

Local dependency

In the community-dwelling setting, evidence of response dependency (person-item residual correlation coefficient >0.3) was observed between 'bathing' and 'dressing' (0.33). Given the borderline correlation, minimal impact on the scale was expected. In the residential care facility setting, no response dependency was observed.

Targeting

The mean location score varied in community-dwelling (Mean: 2.84; SD: 1.62) and residential care facility (Mean: 0.65; SD: 1.85) settings. The positive mean location score in the community-dwelling setting indicates there were insufficient items to discriminate between persons of higher functional ability. This finding is corroborated in Figure 1, Panel A, which visually depicts insufficient overlap between person and item locations. With a mean location score near zero, the scale appeared better targeted for the residential care facility setting (Figure 1, Panel B).

Differential item functioning

Several items exhibited DIF in both settings. Non-uniform DIF by gender and age was clearly evident for 'use of incontinence products' (Figure 2, Panel A, B) and 'financial management' in the community-dwelling setting. Furthermore, non-uniform DIF by gender was observed for 'meal preparation' (Supplemental Figure 1, Panel A).

In the residential care facility setting, 'use of incontinence products' (Figure 2, Panel C) and 'meal preparation' was exhibited DIF by gender (Supplemental Figure 1, Panel B). Notably, there was no significant interaction with class interval, suggesting uniform DIF. 'Getting up from a chair' and 'shopping' exhibited uniform DIF by age whereas 'walking' exhibited non-uniform DIF by age.

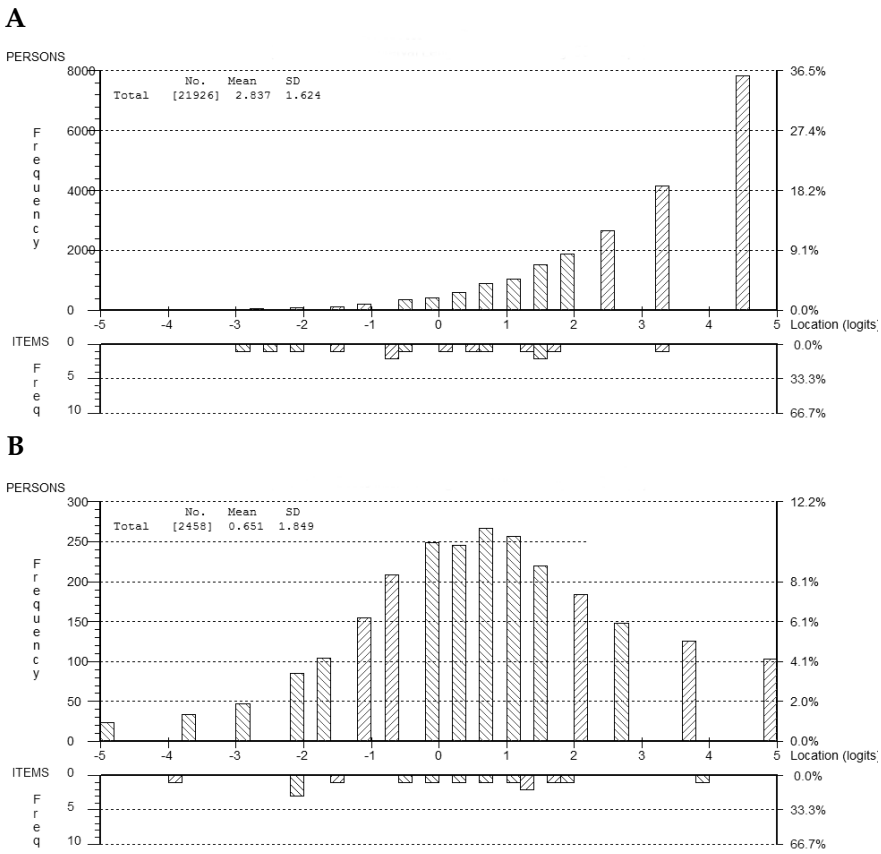


Figure 1 Person-item location distribution maps. (**A**) Person-item location distribution in community-dwelling setting. (**B**) Person-item location distribution in residential care facility setting.

[SD, standard deviation]

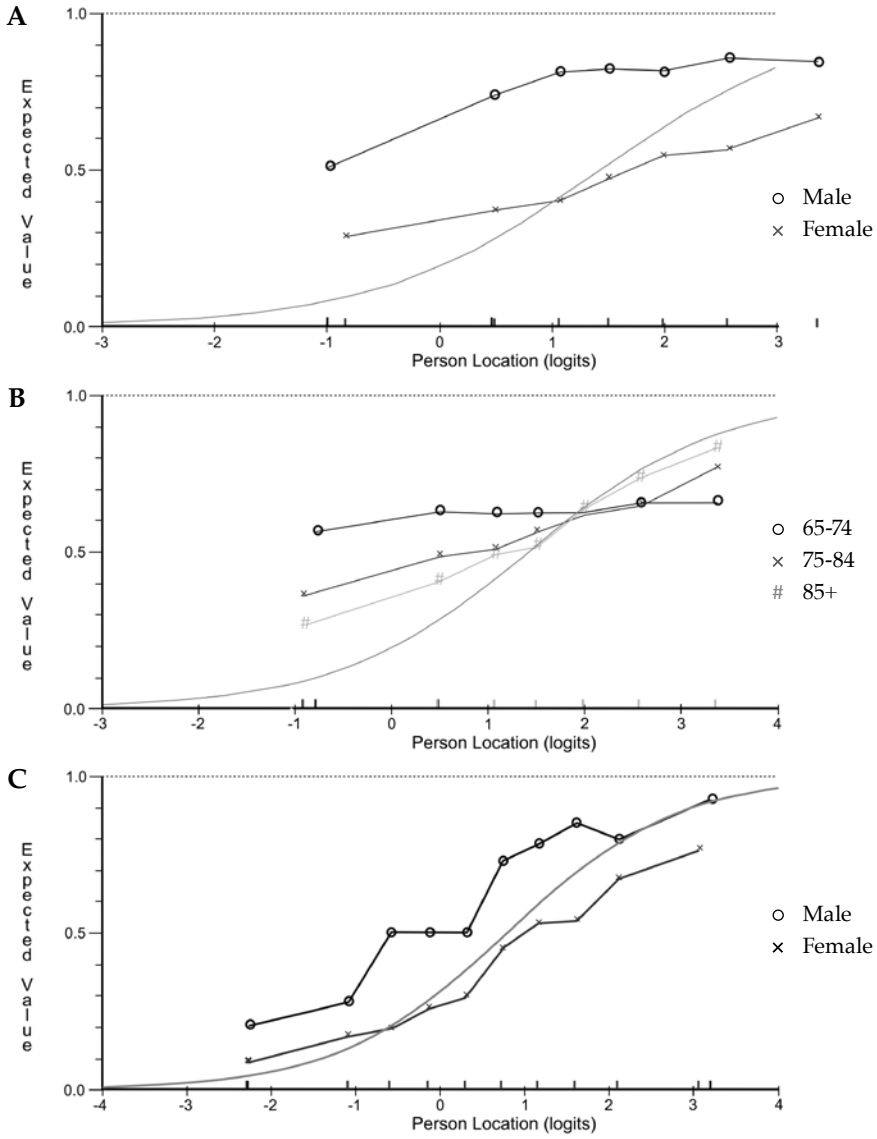


Figure 2 Item characteristic curves (ICC) for ADL item "use of incontinence products" by gender and age group to assess differential information functioning across residential settings. (**A**) ICC Community-dwelling setting by gender, (**B**) ICC Community-dwelling setting by 3 age groups (65-74, 75-84, 85+ years), (**C**) ICC Residential care facility setting by gender.

[ADL, activities of daily living]

Discussion

ADL/IADL scales are often summated to indicate the number of limitations experienced by an individual. However, when performing validity studies, reviewing only the summation of limitations obscures critical information regarding varying difficulty levels between ADL/IADL items. Through applying a Rasch analysis, we found that the 15-item dichotomous ADL/IADL scale exhibited different patterns of difficulty in community-dwelling and residential care facility settings. Our analysis therefore suggests that ADL/IADL summary scores from different residential settings should not be directly compared/combined and that weighting mechanisms should be explored.

This finding is directly relevant for users of TOPICS-MDS, since this large data sharing initiative was designed to allow for the possibility of combining or performing comparative analyses between different study projects. This analysis has demonstrated that even if analyses are stratified by residential setting, the same activities are not necessarily rated similarly across settings. This further impacts health policy suggestions which may be driven by reported levels of limitations. Importantly, this observation is also relevant for researchers who may be interested in comparing ADL/IADL data from different sources. In the era of "Big Data" and data sharing, careful consideration of scale properties are essential prior to performing comparative analyses or merging data.

A second important finding from this study is that the 15-item ADL/IADL scale poorly discriminated between persons with higher ADL/IADL ability in a community-dwelling setting, indicating poor content validity. However, this finding does not necessarily suggest that more items are required in the scale. Rather, discrimination may be improved by providing a wider range of response options.[32]

There are several other notable observations. Whereas the overall fit of the scale to the Rasch model was acceptable in the residential care facility setting, this was not the case for the community-dwelling setting. Speculatively, the poor fit in the community-dwelling sample may be attributed in part to the large sample size or systemic bias in data collection between these two settings. Irrespective of overall fit, at least one individual item exhibited poor fit in both settings. In this regard, further work can be performed to enhance adherence to the Rasch model. Ideally, this work would be conducted during the development phase of the scale when more items can be evaluated to assess fit and discriminative ability.

Despite ADL representing "easier" activities to perform than IADL, we did not anticipate that all six ADL items would be identified as the easiest items. Overlap of ADL and IADL has been previously observed, suggesting that

combined ADL/IADL scales do not follow a rigid hierarchal pattern.[22] Theoretically, however, the ADL scale alone represents a stricter continuum of ability, resulting in the following order of functions from easiest to most difficult: 'eating', 'continence', 'transferring' (e.g. getting up from a chair), 'toileting', 'dressing', 'bathing'. Gerard et al. has previously confirmed this theoretical hierarchy among residents in a skilled nursing facility.[14] Our results from the residential care facility setting demonstrated hierarchal similarities; however, 'use of incontinence products' was rated as more difficult than all other ADL items except for 'bathing'. This difference may be partly attributed to variation in data collection methods. For instance, Gerard et al. extracted ADL data from a pre-existing dataset, and two separate questions (urinary and fecal incontinence) were used to determine use of continence products.[14] In contrast, in TOPICS-MDS, continence was based on a single, dichotomous question.

We observed DIF by gender and age for several items in both settings. Since DIF may indicate that an instrument has poor internal validity, identifying DIF is a critical step in improving measurement and analytical interpretation of results.[6] Non-uniform DIF often requires the removal of the item from the scale[33], whereas weighting is an option for uniform DIF. DIF for ADL/IADL reporting has been previously observed across gender and age groups[17], though some discrepancies for age have been noted.[20] DIF by gender has been partly attributed to historical gender roles. In societies where women have disproportionately performed domestic responsibilities[34], it is unsurprising that differential reporting by gender surfaces in the IADL item, 'meal preparation'. Transitioning away from gender-biased ADL/IADL items is one potential solution in reducing DIF by gender or, if uniform DIF is present, appropriate weightings can also be applied. Underlying mechanisms for DIF by age are not always clear; qualitative analyses can help elucidate why such differences occur.[35] However, generational effects may play a role in reporting differences.

Nonetheless, assessing DIF embodies the overarching aim of achieving invariance in fundamental measurement.[7] Invariance implies that if individuals exhibit similar levels of limitations in activities, responses should not depend on subgroups completing the scale.[36] Further, there is an assumed independence between the overall scores and the set of items being assessed as well as consistency in item difficulty measures across sub-populations.[36] Violation of scale invariance, as observed in our study, jeopardizes group comparisons of ADL/IADL summary scores.

ADL/IADL scales have been routinely used in population surveys for decades. Yet, invariance between subgroups is undesirable for a population surveillance instrument. Adapting ADL/IADL scales for different populations precludes standardization of core measures, and thus modified scales begin to

represent distinct subgroups of activity status.[37] To counter this issue, there has been a movement towards supplementing ADL/IADL scales with other physical measurements, such as 24-hour activity diaries, gait speed and grip strength.[37] It is important to underscore, however, that these measurements are not patient reported outcomes, thereby representing a fundamentally different construct than self-reported activities. Moreover, whereas physical measurements potentially circumvent issues of differential reporting, they have been shown to be influenced by setting, the assessor's level of encouragement and the frequency of the repetition for the activity.[38] Such measurements are also more labor intensive and costly relative to a survey instrument.

Arguably, physical measures may not be required to improve ADL/IADL scales. For instance, the Patient Reported Outcomes Measure Information System (PROMIS) has applied IRT to develop item banks for a range of outcomes, including physical function.[32,39] PROMIS further offers computerized adaptive testing, a method which reduces floor and ceiling effects, quickens assessment time by allowing for the minimum number of relevant scale items and improves measurement precision.[40] The PROMIS physical function item bank confers many advantages when measuring ADL/IADL limitations[40] and is particularly relevant for community-dwelling older persons who may be less likely to require assistance with such activities. Although international validation work is still in progress, initial findings from PROMIS underscore the benefits of IRT in the re-evaluation of instruments, the development of item banks and the use of computer adaptive testing.[32,39]

One limitation of this study is that, to date, there is no universal consensus on the thresholds for testing the unidimensionality of a scale. Whereas this analysis applied established thresholds for acceptable model fit (CFI>0.90; RMSEA<0.08), more conservative thresholds (CFI>0.95; RMSEA<0.06) have been suggested.[41] In the residential care setting, the RMSEA (0.067) did slightly exceed this latter threshold, suggesting a moderate fit. Yet, in line with previous research[18,22], the overall analysis generally supported a unidimensional scale in both settings, a pre-requisite for performing a Rasch analysis.

A second limitation is that the applied ADL/IADL scale may not maintain face validity in other residential care settings internationally. IADL comparative analyses are not always drawn between community-dwelling and residential care settings; however, such analyses maintain relevance in the Netherlands. Unlike nursing homes, in Dutch residential care settings, residents are often encouraged to perform ADL and IADL. Thus, such older persons would not necessarily require professional assistance for 'walking' or many of the IADL included in the scale (e.g. 'grooming', 'traveling', 'using the telephone', 'taking medications', and 'handling finances'). Although 'meal preparation' is often

performed by professional staff, nearly one-third of older persons residing in a residential care setting did not report needing assistance with 'meal preparation' in TOPICS-MDS. This may in part be due to the structure of the question, which does not ask whether older persons *receive* help but rather whether they *need* help. This phrasing was intentional to allow for older persons to report their own perceived ability levels.

Nonetheless, differences in item difficulty persisted even when Rasch analyses were restricted to ADL and walking (results available in Supplementary Tables 1 and 2). Specifically, bathing was rated as an easier activity by older persons residing in community versus residential care settings. It is important to emphasize the reduced number of items resulted in very poor fit to the Rasch model in the community-dwelling setting. Still, this sensitivity analysis further supports the finding of this study that ADL items should not be directly compared between residential settings.

Lastly, although data sharing initiatives provide many opportunities, they also pose analytical challenges. This current analysis was based on 41 research projects with different study aims and protocols (see www.topics-mds.eu for a description of individual research projects). Thus, depending on the research project design, TOPICS-MDS may have been administered as a written questionnaire or an interview. In this analysis, survey mode was mixed in both the community-dwelling (written questionnaire: 59.2%; interview: 40.7%) and the residential care settings (written questionnaire: 28.0%; interview: 72.0%). Survey mode is widely recognized as a potential source of DIF[42], though reassuringly DIF is not necessarily present for all ADL/IADL items.[43] Another potential source of bias is whether older persons received assistance in completing the questionnaire. In the residential care setting, the vast majority of older persons (78.8%) indicated that they themselves chose the response rather than a caregiver.

Yet, the nature of the dataset is also a major strength of this analysis as well. As a large data sharing initiative, TOPICS-MDS is a rich source of information to conduct sub-group analyses. Data sources on older persons are often restricted to certain sub-populations.[17] To the knowledge of the authors, previous research has not applied Rasch analyses to compare ADL/IADL reporting across different residential settings.

Another strength of this analysis is that we applied a Rasch analysis, a well-recognized method to examine the properties of a scale. If model assumptions are met, one of the marked advantages of performing a Rasch analysis is the ability to generate scale-free measures and consistent item difficulties.[36] Importantly, although this type of analysis may be described as producing as sample independent item difficulties, the more appropriate term is

consistent difficulties.[44] Thus, by performing Rasch analysis, we determined the probability of endorsing an ADL or IADL based on individual ability and the item difficulty as measured *on the same scale*. This is in direct contrast to CTT approaches which generate sample dependent results.[36]

In conclusion, participants from different residential settings in TOPICS-MDS exhibited dissimilar hierarchal patterns in rating ADL/IADL difficulty. Scores derived from these settings are not directly comparable and should not be combined together. Moreover, this 15-item dichotomous ADL/IADL scale lacked appropriate content validity to discriminate between older persons with higher ADL/IADL ability. Future adaptations of TOPICS-MDS should consider revising the ADL/IADL domain of the survey.

References

1. Streiner DL, Norman GR. *Health Measurement Scales: A pratical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008.
2. Tractenberg RE. Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemp Clin Trials*. Jan 2010;31(1):1-3.
3. Belvedere SL, de Morton NA. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *J Clin Epidemiol*. Dec 2010;63(12):1287-1297.
4. Tennant A, Conaghan C. The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper? *Arthritis Rheum*. 2007;57(8):1358-1362.
5. Lim SM, Rodger S, Brown T. Using Rasch analysis to establish the construct validity of rehabilitation assessment tools. *Int J Therapy Rehabilitation*. 2009;16(5):251-260.
6. Osterlind SJ, Everson HT. *Differential Item Functioning*. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc. doi: http://dx.doi.org/10.4135/9781412993913; 2009.
7. Krabbe PF. A generalized measurement model to quantify health: the multi-attribute preference response model. *PloS One*. 2013;8(11):e79494.
8. Fusco O, Ferrini A, Santoro M, Lo Monaco MR, Gambassi G, Cesari M. Physical function and perceived quality of life in older persons. *Aging Clin Exp Res*. Feb 2012;24(1):68-73.
9. Holroyd S, Currie LJ, Wooten GF. Depression is associated with impairment of ADL, not motor function in Parkinson disease. *Neurology*. Jun 28 2005;64(12):2134-2135.
10. Laan W, Zuithoff NP, Drubbel I, et al. Validity and reliability of the Katz-15 scale to measure unfavorable health outcomes in community-dwelling older people. *J Nutr Health Aging*. Nov 2014;18(9):848-854.
11. Buurman BM, van Munster BC, Korevaar JC, de Haan RJ, de Rooij SE. Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. *J Clin Epidemiol*. Jun 2011;64(6):619-627.
12. Ward G, Jagger C, Harper W. A review of instrumental ADL assessments for use with elderly people. *Rev Clin Gerontol*. 1998;8(1):65-71.
13. Lindeboom R, Vermeulen M, Holman R, De Haan RJ. Activities of daily living instruments:optimizing scales for neurologic assessments. *Neurology*. Mar 11 2003;60(5):738-742.
14. Gerrard P. The hierarchy of the activities of daily living in the Katz index in residents of skilled nursing facilities. *J Geriatr Phys Ther*. Apr-Jun 2013;36(2):87-91.
15. Kempen GI, Miedema I, Ormel J, Molenaar W. The assessment of disability with the Groningen Activity Restriction Scale. Conceptual framework and psychometric properties. *Soc Sci Med*. Dec 1996;43(11):1601-1610.
16. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc*. Jun 1992;40(6):585-588.
17. Fieo RA, Austin EJ, Starr JM, Deary IJ. Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into the preclinical range: a systematic review. *BMC Geriatr*. 2011;11:42.
18. das Nair R, Moreton BJ, Lincoln NB. Rasch analysis of the Nottingham extended activities of daily living scale. *J Rehabil Med*. Oct 2011;43(10):944-950.
19. Lutomski JE, Baars MA, Schalk BW, et al. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PloS One*. 2013;8(12):e81673.
20. LaPlante MP. The classic measure of disability in activities of daily living is biased by age but an expanded IADL/ADL measure is not. *J Gerontol B Psychol Sci Soc Sci*. Nov 2010;65(6):720-732.
21. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. J *Gerontol B Psychol Sci Soc Sci*. Sep 2002;57(5):S275-284.

22. Spector WD, Fleishman JA. Combining activities of daily living with instrumental activities of daily living to measure functional disability. *J Gerontol B Psychol Sci Soc Sci*. Jan 1998;53(1):S46-S57.
23. Zucca A, Lambert SD, Boyes AW, Pallant JF. Rasch analysis of the Mini-Mental Adjustment to Cancer Scale (mini-MAC) among a heterogeneous sample of long-term cancer survivors: a cross-sectional study. *Health Qual Life Outcomes*. 2012;10:55.
24. van der Wal MB, Tuinebreijer WE, Lundgren-Nilsson A, Middelkoop E, van Zuijlen PP. Differential item functioning in the Observer Scale of the POSAS for different scar types. *Qual Life Res*. Sep 2014;23(7):2037-2045.
25. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Software*. 2012;48(2):1-36.
26. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Soc Methods Res*. 1993;21:230-258.
27. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. Mar 1990;107(2):238-246.
28. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6:1-55.
29. Tennant A, Pallant JF. The root mean square error of approximation (RMSEA) as a supplementary statistic to determine fit to the Rasch model with large sample sizes. *Rasch Measurement Transactions*. 2012;25(4):1348-1349.
30. Wright BD. Logits? *Rasch Measurement Transactions*. 1993;7(2):288.
31. Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry*. 2009;9:21.
32. Bruce B, Fries JF, Ambrosini D, et al. Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther*. 2009;11(6):R191.
33. Pallant JF, Miller RL, Tennant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry*. 2006;6:28.
34. Treas J, Lui J. Studying housework across nations. *J Family Theory Rev*. 2013;5(2):135-149.
35. Paz SH, Spritzer KL, Morales LS, Hays RD. Age-related Differential Item Functioning for the Patient-Reported Outcomes Information System (PROMIS(R)) Physical Functioning Items. *Prim Health Care*. Mar 29 2013;3(131).
36. Engelhard Jr G. *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. New York, NY, USA: Routledge; 2013.
37. National Research Council. *Improving the Measurement of Late-Life Disability in Population Surveys: Beyond ADLs and IADLs: Summary of a Workshop*. Washington, DC, USA: The National Academies Press; 2009.
38. Guralnick JM, Branch LG, Cummings SR, JD. C. Physical performance measures in aging research. *J Gerontol: Med Sci*. 1989;44(5):M141-146.
39. Aletaha D. From the item to the outcome: the promising prospects of PROMIS. *Arthritis Res Ther*. 2010;12(1):104.
40. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE, Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol*. May 2014;67(5):516-526.
41. Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res*. May 2009;18(4):447-460.
42. Martin P, Lynn P. The effects of mixed mode survey designs on simple and complex analyses; Centre for Comparative Social Surveys Working Paper Series: Paper no. 04. Centre for Comparative Social Surveys, City University London2011.
43. Lutomski JE, Krabbe PFM, Bleijenberg N, et al. ADL and IADL reporting are biased by survey mode. *J Am Geriatr Soc*. 2015;In press.
44. van der Linden W. Sample Independency in the Rasch Model? *Rasch Measurement Transactions*. 1993;6(4):247.

Supplementary Table 1 Overall scale and person fit to the Rasch Model based on six ADL items + mobility

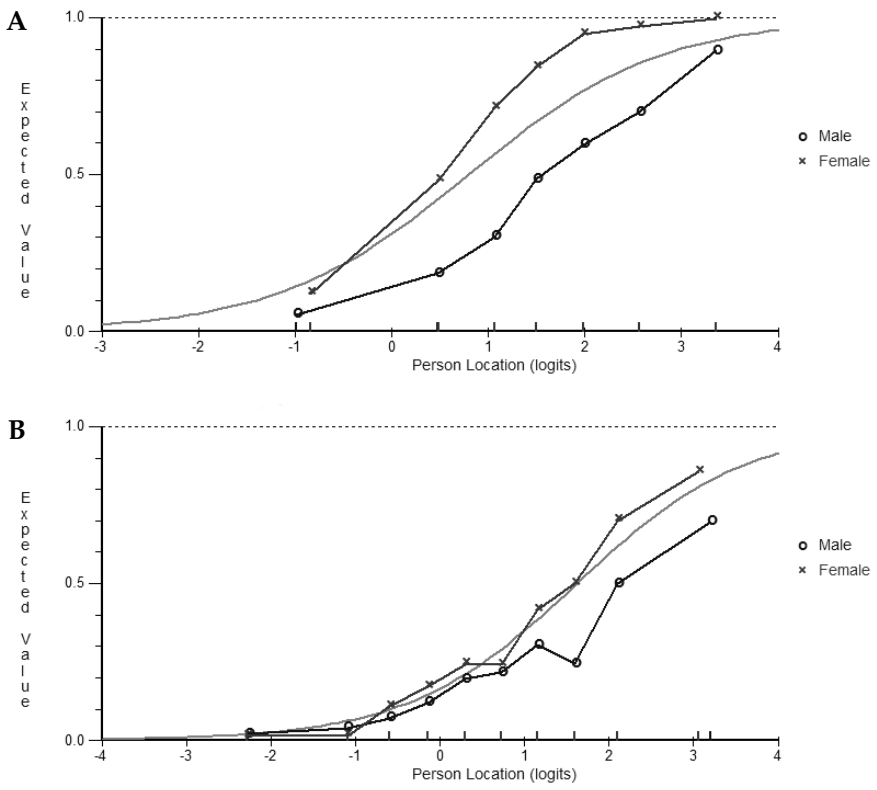
Residence type	Standardized mean item fit residual	SD	χ^2	p-value	RMSEA	Mean standardized person fit residual	SD	PSI with extremes	PSI without extremes
Community-dwelling	-4.54	18.77	3,895.68	<0.0001	0.09	-0.24	0.48	-0.25	0.08
Residential care facility	-1.29	4.69	609.65	<0.0001	0.08	-0.28	0.51	0.62	0.53

Abbreviations: ADL, activities of daily living; sd, standard deviation; RMSEA, root mean square error of approximation; PSI, person separation index
Note: A mean fit residual of 0 with a standard deviation between 1 and -1 suggest perfect fit. Whereas a non-significant χ^2 value typically indicates good fit, in large samples, a RMSEA value <0.2 indicates adequate model. A PSI score >0.70 is recommended for scales interpreted at the population level.

Supplementary Table 2 Item fit statistics for six ADL items + mobility in community-dwelling setting (N=21,926) and residential care facility setting (N=2,458), ordered by item location

Item	Location	SE	Fit residual	χ^2	p-value
Community-dwelling					
Eating	-2.49	0.06	-1.85	64.14	<0.0001*
Toileting	-1.88	0.05	-13.06	188.99	<0.0001*
Getting up from chair	-0.22	0.03	-10.85	232.50	<0.0001*
Dressing	0.05	0.03	-21.46	658.17	<0.0001*
Bathing	0.75	0.03	-20.49	848.46	<0.0001*
Walking	1.87	0.02	32.81	1636.74	<0.0001*
Use of incontinence products	1.91	0.02	3.09	266.68	<0.0001*
Residential care facility					
Eating	-3.35	0.11	-0.43	44.83	<0.0001*
Toileting	-1.75	0.08	-5.39	109.52	<0.0001*
Getting up from chair	-0.94	0.07	-2.22	60.85	<0.0001*
Dressing	0.57	0.06	-7.20	162.76	<0.0001*
Use of incontinence products	1.50	0.05	7.18	119.40	<0.0001*
Walking	1.93	0.06	1.22	13.24	0.02
Bathing	2.05	0.06	-2.19	99.04	<0.0001*

Abbreviation: ADL, activities of daily living; SE, standard error
Note: *Bonferroni adjusted significant level, 0.000667



Supplemental Figure 1 Item characteristic curves (ICC) for IADL item 'meal preparation by gender to assess differential information functioning across residential settings. (**A**) ICC Community-dwelling setting by gender, (**B**) ICC Residential care facility setting by gender.

[IADL, instrumental activities of daily living]

3.2

The influence of survey mode on ADL and IADL reporting

Accepted for publication as:

JE Lutomski, Krabbe PFM, Bleijenberg N, Blom JW, Buurman BM, Kempen GIJM, Muntinga ME, Steyerberg EW, Olde Rikkert MGM, Melis RJF. ADL and IADL reporting are biased by survey mode. *Journal of the American Geriatrics Society*. 2015; 63(11):2419-21.

To the Editor,

Disparities between outcome data generated from interviews versus written questionnaires is a well-recognized phenomenon in health research.[1] Such disparities may be due to various reasons, such as the drive to give socially desirable answers or the fact that respondents may interpret written and spoken questions differently.[2] In aging research, scales measuring Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) are no exception, with written questionnaires often increasing the likelihood of reported ADL and IADL independence.[3]

Reporting disparities, also referred to as "differential item functioning" (DIF), represent a type of measurement bias whereby participant subgroups (e.g. those completing an interview versus a written questionnaire) have different probabilities of reporting an outcome (e.g. an ADL/IADL item) even when they possess the same ability level.[4] To examine DIF across survey modes, researchers frequently perform a comparative analysis against a reference standard, such as clinical assessment. However, in the absence of clinical assessment, alternative statistical techniques, such as Rasch analysis, can demonstrate DIF.[4] The purpose of our study was to assess DIF in self-reported ADL and IADL by survey mode.

Methods

Data were derived from The Older Persons and Informal Caregivers Minimum DataSet (TOPICS-MDS; www.topics-mds.eu), which is a public access dataset designed to capture essential health and wellbeing information on older persons and informal caregivers in the Netherlands.[5] Briefly, TOPICS-MDS consists of prospectively collected data and includes a modified version of the Katz Index of Independence of Activities in Daily Living[6], comprising of 15 dichotomous ADL and IADL items. Our analysis is based on 1,937 participants aged ≥65 years residing in residential care facilities in the Netherlands with complete data on ADL/IADL functioning and survey mode.

To assess DIF by survey mode, we applied a Rasch measurement model, which assumes that the probability of item endorsement is a logistic function of the relative distance between the item's difficulty and the participant's ability.[7] In other words, persons with similar ability levels are more likely to report to assistance with similar ADL/IADL (e.g. persons requiring assistance with eating also likely require assistance with dressing). Based on this model, we performed an ANOVA of the standardized response residuals of each ADL/IADL item; significant differences suggest DIF.[7,8] We visually confirmed DIF by

inspecting item characteristic curves by survey mode.[7,8] These curves represent the probability of reporting an ADL/IADL item based on the ability level of the participant completing the scale. If these curves follow different trajectories, DIF is present. Analyses were conducted using RUMM2030 (Rumm Laboratory, Perth, Western Australia).

Results

Of the 1,937 participants, 1,421 (73.4%) were women. The average age was 85 years (±6 years), ranging from 65 to 102 years. Overall, 1,395 (72.0%) were interviewed whereas 542 (28.0%) completed a written questionnaire. Participants were most likely to require assistance with household tasks (n=1,729; 89.3%) and least likely with eating (n=72; 3.7%).

Several ADL/IADL items clearly exhibited DIF by survey mode: use of incontinence products, walking and meal preparation (ANOVA results not shown). Participants who were interviewed were more likely to report using incontinence products than those completing a written questionnaire (Figure 1, Panel A). Among participants with lower functional ability (i.e. <0 on the logit scale), the probability of reporting assistance with walking was higher among those who completed a written questionnaire. This disparity diminished among participants with higher functional ability (i.e. >0 on the logit scale; Figure 1, Panel B). The converse pattern was observed for meal preparation (Figure 1, Panel C).

Discussion

DIF is problematic for health measurement scales, and if serious, requires the exclusion of an item. Reassuringly, DIF by survey mode was not present for all ADL/IADL items. Nonetheless, meal preparation exhibited DIF by survey mode in this study, as well by gender in previous research[9]; removing this item from the scale may be warranted. We are unable to confirm why DIF occurred for use of incontinence and walking, though speculatively the interaction with the interviewer may have prompted contrasting response patterns in participants with different functional levels. Researchers interested in the construction or refinement of ADL/IADL scales should consider Rasch analysis to identify items prone to reporting disparities to guide item selection.

Our study demonstrates that reporting disparities by survey mode are prominent for certain ADL/IADL items and supports a single data collection

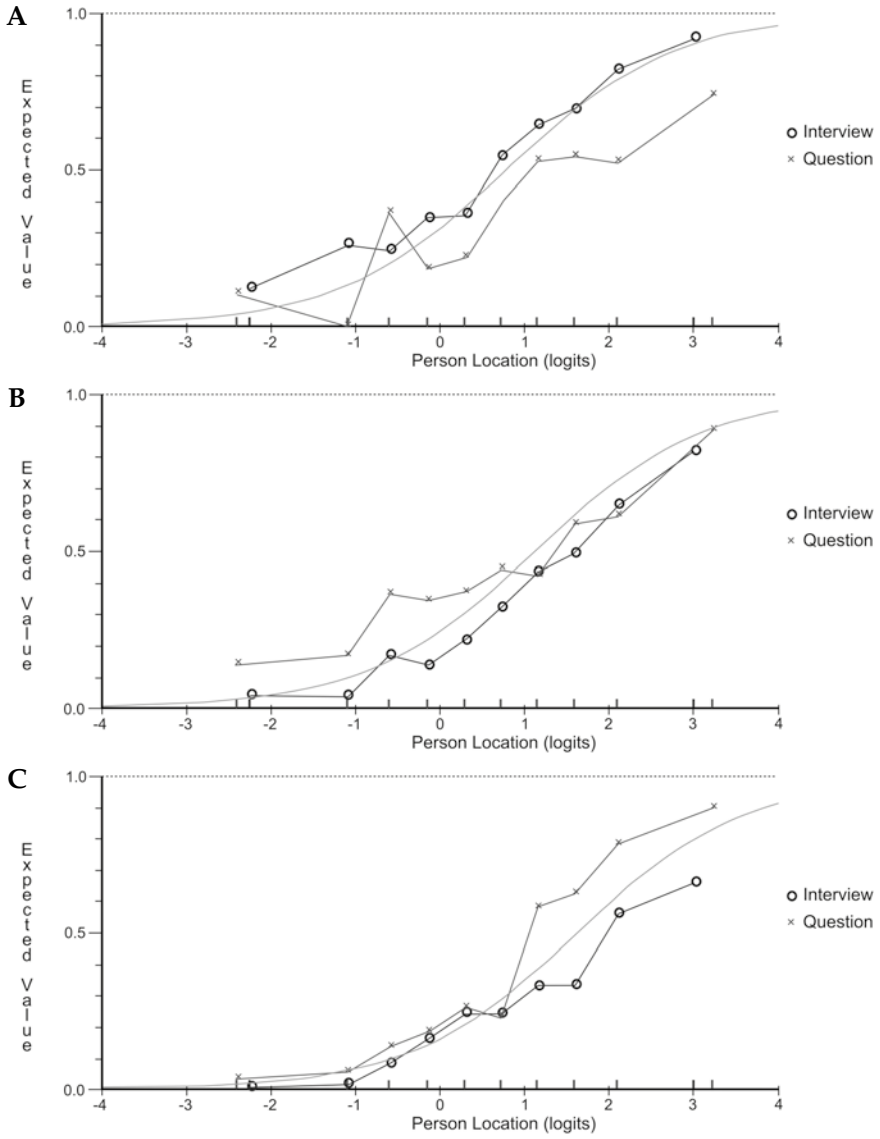


Figure 1 Item characteristic curves for select activity of daily living (ADL) or instrumental activity of daily living (IADL) ((**A**) use of incontinence products, (**B**) walking, (**C**) meal preparation) according to survey mode.

The x-axis (person location) represents participants' underlying ability based on a standardized logit scale, ranging from less functional ability (negative values) to greater functional ability (positive values). The y-axis (expected value, possible range 0–1) refers to the probability that participants reported requiring assistance with a specific ADL or IADL.

mode.[2] However, in aging populations with varying abilities to complete a survey, a single data collection mode may not always be realistic. Careful consideration of how to treat mixed mode data should be undertaken in the analytical phase and may require the use of subgroup analyses or interaction terms.[2]

References

1. Hood K, Robling M, Ingledew D, et al. Mode of data elicitation, acquisition and response to surveys: a systematic review. *Health Technol Assess*. May 2012;16(27):1-162.
2. Martin P, Lynn P. The effects of mixed mode survey designs on simple and complex analyses; Centre for Comparative Social Surveys Working Paper Series: Paper no. 04. Centre for Comparative Social Surveys, City University London; 2011.
3. Waehrens EE, Bliddal H, Danneskiold-Samsoe B, Lund H, Fisher AG. Differences between questionnaire- and interview-based measures of activities of daily living (ADL) ability and their association with observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis, and fibromyalgia. *Scand J Rheumatol*. Mar 2012;41(2):95-102.
4. Osterlind SJ, Everson HT. *Differential Item Functioning*. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc. doi: http://dx.doi.org/10.4135/9781412993913; 2009.
5. Lutomski JE, Baars MA, Schalk BW, et al. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PloS One*. 2013;8(12):e81673.
6. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc*. Jun 1992;40(6):585-588.
7. Tennant A, Conaghan C. The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper? *Arthritis Rheum*. 2007;57(8):1358-1362.
8. Pallant JF, Miller RL, Tennant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry*. 2006;6:28.
9. McDowell I. Physical disability and handicap. *Measuring health: A guide to rating scales and questionnaires*. New York, New York, USA: Oxford University Press, Inc.; 2006:55-149.

3.3

The performance of the EQ-5D across geriatric conditions

Submitted for publication as:

Lutomski JE, Krabbe PFM, Bleijenberg N, Blom J, Kempen GIJM, MacNeil-Vroomen J, Muntinga ME, Steyerberg EW, Olde-Rikkert MGM, Melis RJF, on behalf of TOPICS Consortium. Measurement properties of the EQ-5D across four major geriatric conditions: Findings from TOPICS-MDS.

Abstract

Background: As populations age, chronic geriatric conditions linked to progressive organ failure jeopardize health-related quality of life (HRQoL). Since the EQ-5D is one of the commonest instruments used to measure HRQoL, this study assessed the validity and applicability of the EQ-5D in four major chronic geriatric conditions: hearing issues, joint damage, urinary incontinence, or dizziness with falls. Outcomes were compared to the overall study population as well as a 'healthy' subgroup reporting no major chronic conditions.

Methods: Data on 25,637 community-dwelling persons aged 65 years and older were derived from the Dutch public access data repository, The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS, www.topics-mds.eu). Floor and ceiling effects were examined. To assess convergent validity, random effects meta-correlations (Spearman's rho) were derived between individual EQ-5D domains and related survey items. To further examine construct validity, the association between sociodemographic characteristics and EQ-5D summary scores were assessed using linear mixed models.

Results: Whereas ceiling effects were observed in the overall study population and the 'healthy' subgroup, such was not the case in geriatric condition subgroups. Most hypotheses regarding associations between EQ-5D item scores and related constructs and between EQ-5D summary scores and sociodemographic characteristics were supported. Strong correlations were observed between the EQ-5D self-care item and the Katz Index bathing and dressing items as well as between the EQ-5D anxiety/depression item and the Rand-36 mental health sub-scale summary score. Moderate correlations were observed between the EQ-5D summary and general quality of life scores. EQ-5D summary scores were lower in respondents who were older, female, widowed/single, lower educated, and living alone. Increasing co-morbidity had a clear negative effect on EQ-5D scores.

Conclusion: This study supported the construct validity of the EQ-5D across four major geriatric conditions. For older persons who are generally healthy, i.e. reporting few to no chronic conditions, the EQ-5D confers poor discriminative ability due to ceiling effects. Although the overall dataset initially suggested poor discriminative ability for the EQ-5D, such was not the case within subgroups presenting with major geriatric conditions.

Introduction

Medicine is transitioning away from the traditional single disease model which often dichotomizes health as the presence or absence of disease. Rather, to encompass the far-reaching effects of morbidity, global measures of health, such as health-related quality of life (HRQoL), are now viewed as critical outcomes in contemporary medical research.[1] HRQoL is a subjective, multidimensional concept which incorporates physical, psychological, and social wellbeing. Numerous instruments have been designed to measure HRQoL.[2] Some subgroups of individuals, such as cancer patients[3], have their own HRQoL instruments whereas others are intended for general use.[2] The EQ-5D falls under this latter category and is one of the commonest instruments used to measure HRQoL. The EQ-5D was developed by an international task force to permit a quick assessment of an individual's overall health status across five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[4] The EQ-5D can be used to generate a descriptive profile as well as a preference-weighted summary score.[4]

Although the broad scope and simplicity of the EQ-5D have made it a widely accessible instrument, these characteristics have also elicited ceiling effects in data collection.[5-7] Thus, a prevailing concern is that the relatively few dimensions may lack the necessary level of responsiveness and discriminative ability to detect health changes for certain health conditions.[8] Numerous studies have therefore investigated measurement properties of the EQ-5D across a wide range of morbidity subgroups.[7,9,10] Although construct validity is usually maintained, there have been observed differences in the instrument's performance. For instance, the EQ-5D was more responsive in persons with urinary incontinence [11] than in persons with hearing impairment.[12]

Still, there remains a fundamental need to re-evaluate measurement properties of the EQ-5D in different study populations. Although previous validation work gives credence to the robustness of EQ-5D, the instrument itself is never truly 'valid'. Rather, validity is a characteristic of a study population[13]; measurement properties are thus intrinsically linked to the target sample (e.g. community-dwelling older persons), the country of the study, how subgroups of interest were identified, and other defining characteristics of the study population. For this reason, preliminary validation of the instrument is necessary prior to in-depth analyses. Moreover, in older persons' research, there is also a clear benefit to focus on the instrument's performance across geriatric conditions, i.e. acute or chronic conditions in older persons stemming from multifactorial causes.[14] As populations age, chronic geriatric conditions linked to progressive organ failure are of particular interest since they place a growing demand

on health care services and further jeopardize health-related quality of life (HRQoL).

Since the EQ-5D is one of the commonest instruments used to measure HRQoL, the primary aim of this study was to assess the validity and applicability of the EQ-5D in four major chronic geriatric conditions: hearing issues, joint damage, urinary incontinence, or dizziness with falls. These conditions were among the most frequently reported conditions in the dataset under review and in general are prevalent in older populations. Whereas a substantial number of studies have examined singular conditions, we complement previous research by further evaluating the measurement properties of these subgroups against the overall study population as well as a 'healthy' subgroup reporting no major chronic conditions. For the purposes of this study, we assessed floor and ceiling effects and construct validity.

Methods

Data source

Data were derived from The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS), which is a public access data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere (www.topics-mds.eu).[15] Briefly, the Dutch National Care for the Elderly Programme (http://www.nationaalprogrammaouderenzorg.nl) was established in 2008 to promote proactive, integrated health care for older persons with complex care needs. As part of this national agenda, TOPICS-MDS was developed to prospectively collect uniform information from all research projects funded under the programme. Thus, TOPICS-MDS consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardized protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to TOPICS-MDS, the database is dynamic in nature and may be updated with new observations. Our present analysis is based on the second version of the database (available as of May 2015). TOPICS-MDS is a fully anonymized dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Centre Ethical Committee review reference number: CMO: 2012/120).

Population

TOPICS-MDS includes studies which sampled from institutionalized (nursing homes and residential care facilities) and non-institutionalized settings. For the purposes of this study, analyses were based on community-dwelling older persons aged 65 years and older. Since sampling strategies varied across study protocols, older persons were sampled from primary care centers, hospital settings, or the general population.

Variables

The EQ-5D assesses five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) using a three-level response option (1=no problems; 2=some problems; 3=extreme problems).[16] When combined, these scores can describe up to 243 (i.e. 3^5) unique health states, with '11111' and '33333' representing the best and worst possible health states respectively. A summary score can also be derived using a population-specific tariff (weighting); this analysis has applied a tariff validated for the Dutch population.[17] An EQ-5D summary score of one represents the best imaginable health status whereas a score lower than zero represents a health status perceived to be worse than death.

Morbidity status was self-reported. For the purposes of this study, subgroup analyses focused on the four most prevalent geriatric conditions in the database. Older persons were asked if they had experienced hearing issues, joint damage (defined as arthrosis or degenerative arthritis of the hips or knees), urinary incontinence, or dizziness with falls in the past 12 months. In TOPICS-MDS, many respondents reported more than one condition; thus, subgroups were not mutually exclusive. Thirteen other chronic conditions which are regularly recorded in the older Dutch population were also assessed.[18] Respondents were classified as 'healthy' if they did not report any of the 17 conditions assessed in TOPICS-MDS.

There were several variables of interest for validation purposes. Limitations in activity were determined using a modified 15-item Katz Index[19,20], which included Activities of Daily Living (ADL; i.e. bathing, dressing, toileting, use of incontinence products, transferring, and eating), Instrumental Activities of Daily Living (IADL; i.e. grooming, use of a telephone, travelling, grocery shopping, meal preparation, household tasks, taking medication, and financial management) and an additional indicator for mobility (i.e. walking). Response options were dichotomized as 'requires assistance' or 'does not require assistance'. An IADL summary score was summated, ranging from zero to eight, with higher scores representing greater limitations in activities. Emotional wellbeing was evaluated using the Rand-36 mental health sub-scale.[21] This scale asked how often in the past four weeks an individual has felt: very nervous;

calm and peaceful; down-hearted and blue; happy; so down in the dumps nothing could cheer [him/her] up. A five-level response option was presented ranging from 'never' to 'always'. Positive items were scored from zero to 100 whereas negative items were reverse scored. A summary score ranged zero and 100 with higher scores implying a more positive emotional state. Self-perceived general quality of life was assessed with a modified version of Cantril's Self Anchoring Ladder.[22] Older persons were asked to rate their present life on a scale between zero (completely unsatisfied with life) and ten (completely satisfied with life).

Statistical analysis

Given that TOPICS-MDS is a pooled dataset, subsequent analyses were derived using a one-step individual patient data meta-analysis.[23] Demographic characteristics of the study population were assessed. Distributional properties were derived for individual EQ-5D items. The mean (standard deviation), range and floor and ceiling effects for the EQ-5D summary score were further derived. To date, there is no general consensus for floor and ceiling effects; thus, these effects were considered to be present if at least 15% of older persons reported either the lowest scores (health state '33333', i.e. weighted score -0.33) or highest scores (health state '11111', i.e. weighted score 1.0).[24]

To assess convergent validity, random effects meta-correlations (Spearman's rho) were derived using the meta package in R to allow for heterogeneity between individual studies in the pooled dataset.[25] The EQ-5D mobility item was correlated with Katz Index item, 'assistance with walking'. The EQ-5D self-care item was correlated with two ADL items from the Katz Index, bathing and dressing. The EQ-5D usual activities item was correlated with the summary IADL score. The anxiety/depression item was correlated with the Rand-36 mental health sub-scale summary score. Lastly, the EQ-5D summary score was correlated with a general quality of life score (Cantril's Self Anchoring Ladder). Correlation coefficients were classified as trivial ($\leq$0.1), weak (0.1 to <0.3), moderate (0.3 to <0.5), strong (0.5 to <0.7), very strong ($\geq$0.7).[26] A strong to very strong, positive correlation was hypothesized between the EQ-5D mobility, self-care, usual activities items and the Katz walking item, the Katz bathing/dressing items and the Katz IADL summary score respectively. A moderate to strong, negative correlation was anticipated between the EQ-5D anxiety/depression item and the Rand-36 mental health sub-scale summary score. The general quality of life score based on Cantril's Self Anchoring Ladder is a broad evaluative measure whereas the EQ-5D summary score reflects HRQoL. Given this conceptual distinction, a moderate positive correlation was hypothesized between these scores.

To further examine construct validity and to allow for clustering effects by individual research projects, linear mixed models were performed by regressing the EQ-5D summary score on key demographic variables in the overall study population as well as within subgroups. Variables included age, sex, marital status, educational level, living arrangement (i.e. alone or with others), and morbidity status. HRQoL was hypothesized to be lower in respondents who were older[27], widowed or single (defined as unmarried or divorced)[28,29], lower educated[27], and living alone.[30] Moreover, women were expected to report lower overall HRQoL[29], and in particular to report higher anxiety and/or depression.[27] Multimorbidity was anticipated to have a strong, negative effect on HRQoL.[31,32] Associations were examined in unadjusted models as well as models adjusted for age and sex. All statistical analyses were carried out using SPSS (Version 21.0. Armonk, NY, USA: IBM Corp) and R (2013: Vienna, Austria).

Results

Data were extracted on 25,637 community-dwelling persons aged 65 years and older from 32 research projects. Most were sampled from primary care centers (69.3%, n=17,777), followed by hospital settings (20.6%, n=5,294) and the general population (10.0%, n=2,566). From the overall study population, nearly three-quarters (73.3%, n=18,791) reported at least one of the four geriatric conditions under review and 8.9% (n=2,275) were classified as 'healthy', i.e. reported no morbidities.

Sociodemographic characteristics differed across subgroups (Table 1). Respondents comprising the 'healthy' subgroup were more likely to be younger, married, higher educated, and reside with others. These respondents reported the highest psychological wellbeing and general quality of life scores. Whereas the average age was broadly similar across geriatric condition subgroups, there were observable differences in the distributions of sex and marital status. Nearly half of respondents in the hearing issues subgroup were male compared to one-quarter in the urinary incontinence subgroup. Respondents in the urinary incontinence group were the most likely to be widowed. Based on the Katz Index, assistance needed for bathing was disproportionately higher in the urinary incontinence and dizziness with falls subgroups; the median number of reports limitations in IADL was also higher in these subgroups. Respondents in the dizziness with falls subgroup reported the lowest scores for psychological wellbeing and general quality of life.

In the overall study population, there was a clear ceiling effect, with nearly one in five respondents (19.2%) reporting optimal HRQoL (i.e. an EQ-5D score of

'11111') (Table 2). This effect was driven, in part, by respondents in the 'healthy' subgroup. These respondents were the least likely to report any problems across the five EQ-5D dimensions, and more than half (57.5%) reported optimal HRQoL.

Table 1 Sociodemographic characteristics of community-dwelling older persons aged 65 years and older, TOPICS-MDS, 2015

	Overall sample N=25,637	Healthy[a] N=2,275	Hearing issues N=9,762	Joint damage[b] N=11,903	Urinary incontinence N=5,932	Dizziness with falls N=4,273
Age [mean (SD)]	78 (6)	75 (5)	80 (6)	78 (6)	79 (6)	79 (6)
Sex						
Men	41.7	49.2	46.1	29.9	25.0	34.9
Women	58.3	50.8	53.9	70.1	75.0	65.1
Marital status						
Married/cohabiting	53.9	63.7	51.8	48.0	42.9	44.7
Widowed	35.6	26.5	38.8	40.8	45.9	44.0
Single/divorced	10.5	9.8	9.3	11.2	11.2	11.3
Educational level						
Primary	31.7	24.1	33.1	34.5	37.2	37.8
Secondary	49.2	51.6	47.6	48.9	47.5	46.2
College/Some college	19.1	24.3	19.3	16.6	15.4	16.0
Living arrangements						
Lives alone	44.9	34.7	46.8	50.8	54.9	54.0
Lives with others	55.1	65.3	53.2	49.2	45.1	46.0
Morbidity						
Hearing issues	39.3	--	100	42.2	47.0	51.0
Joint damage	48.5	--	52.2	100	61.9	59.8
Urinary incontinence	24.0	--	29.0	31.1	100	37.4
Dizziness with falls	17.3	--	22.5	21.3	26.6	100
Other co-morbidities[c]						
None	21.7	--	15.7	16.2	12.2	10.6
One	31.0	--	28.4	28.6	25.8	23.2
Two	24.0	--	26.1	24.8	25.8	25.6
Three or more	23.4	--	29.8	30.3	36.3	40.7

Table 1 Continued

	Overall sample N=25,637	Healthy[a] N=2,275	Hearing issues N=9,762	Joint damage[b] N=11,903	Urinary incontinence N=5,932	Dizziness with falls N=4,273
Requires assistance[d]						
Bathing	15.6	4.1	17.0	19.0	25.4	25.4
Dressing	11.6	3.0	2.6	2.8	4.1	4.3
Walking	25.8	6.8	29.8	34.6	39.9	42.8
IADL score[d] [median (IQR)]	1 (2)	0 (0)	1 (3)	1 (3)	2 (2)	2 (2)
Psychological wellbeing[e] [mean (SD)]	74.2 (18.0)	83.5 (14.0)	72.8 (17.9)	71.3 (18.3)	69.3 (18.7)	65.9 (19.5)
General quality of life[f] [mean (SD)]	7.4 (1.3)	8.0 (1.0)	7.2 (1.3)	7.2 (1.3)	7.0 (1.3)	6.9 (1.4)

Note: Percentages are presented unless other specified. **Abbreviations:** IADL, Instrumental Activities of Daily Living; IQR, interquartile range.

[a] Reported having none of the 17 recorded chronic conditions.

[b] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

[c] Included self-reported diabetes, cerebrovascular events, heart failure, cancer, airway disease, osteoporosis, bone fractures, depression, anxiety, dementia, or vision problems.

[d] Modified Katz Index of Independence in Activities of Daily Living. IADL score ranges from zero to eight limitations.

[e] Rand mental health subscale (Range 0-100; higher scores represent a more positive emotional state).

[f] Cantril's Self Anchoring Ladder (Range 0-10; higher scores represent higher perceived quality of life).

Relative to the joint damage, urinary incontinence and dizziness with falls subgroups, respondents in the hearing issues subgroup were the least likely to report any problems across the five dimensions. Yet, despite a higher proportion of respondents in the hearing issues subgroup reporting optimal HRQoL, a ceiling effect was not observed. Irrespective of the subgroup, very few respondents (<1%) reported the worst imaginable health state (i.e. an EQ-5D score of '33333').

Differences in population size and reporting levels across individual dimensions of the EQ-5D attributed to differences in the number of observed health states (defined as the concatenation of domain scores). In the overall study population, 213 out of the 243 potential health states were represented. One hundred ninety-one profiles were represented in the hearing issues subgroup, 194 in the joint damage subgroup, 190 in the urinary incontinence subgroup, and 169 in the dizziness with falls subgroup. In the 'healthy'

Table 2 Distributional properties of the EQ-5D summary score and individual dimensions

	Overall sample N=25,637	Healthy[a] N=2,275	Hearing issues N=9,762	Joint damage[b] N=11,903	Urinary incontinence N=5,932	Dizziness with falls N=4,273
EQ-5D dimensions						
Mobility						
No problems	39.0	77.4	33.1	22.5	22.0	23.2
Slight problems	57.4	21.8	63.8	73.1	72.7	73.1
Extreme problems	3.6	0.8	3.0	4.4	5.2	3.7
Self-care						
No problems	79.2	94.9	76.8	73.6	67.2	67.2
Slight problems	15.7	3.8	18.1	20.4	23.6	24.4
Extreme problems	5.1	1.3	5.1	6.0	9.2	8.4
Usual activities						
No problems	56.4	86.8	52.3	45.2	41.5	41.4
Slight problems	34.2	10.2	38.4	43.6	45.2	45.0
Extreme problems	9.3	3.0	9.3	11.2	13.3	13.5
Pain/Discomfort						
No problems	37.4	71.2	34.3	20.3	25.2	23.6
Slight problems	53.5	27.0	55.7	66.1	60.3	60.1
Extreme problems	9.2	1.8	10.0	13.6	14.5	16.2
Anxiety/Depression						
No problems	75.3	91.3	72.7	70.7	65.0	60.3
Slight problems	22.4	8.4	25.0	26.5	30.8	34.3
Extreme problems	2.3	0.3	2.4	2.8	4.2	5.4
EQ-5D summary score						
Mean (SD)	0.72 (0.26)	0.90 (0.16)	0.71 (0.26)	0.65 (0.26)	0.63 (0.28)	0.61 (0.29)
Range	-0.33, 1.00	-0.18, 1.00	-0.33, 1.00	-0.33, 1.00	-0.33, 1.00	-0.33, 1.00
Floor (%)	<1.0	0	<1.0	<1.0	<1.0	<1.0
Ceiling (%)	19.2	57.5	14.1	6.3	7.3	6.8

Note: Percentages are presented unless other specified
[a] Reported having none of the 17 recorded chronic conditions recorded.
[b] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

subgroup, 76 different health profiles were observed. Optimal HRQoL was the most frequently reported health state in the overall study population (19.2%) as well as in the 'healthy' (57.5%) and hearing issues (14.9%) subgroups (Table 3). In contrast, the most frequently reported profile in the joint damage subgroup was "some issues" with mobility and pain/discomfort and "no issues" with self-care, usual activities or anxiety/depression (13.8%). The urinary incontinence and dizziness with falls subgroups mirrored the joint damage subgroup, with the exception of reporting "some issues" for usual activities (10.5% and 9.4% respectively).

Table 3 Ten most frequently reported EQ-5D health states[a]

Overall sample N=25,637 Profile (%)	Healthy[b] N=2,275 Profile (%)	Hearing issues N=9,762 Profile (%)	Joint damage[c] N=11,903 Profile (%)	Urinary incontinence N=5,932 Profile (%)	Dizziness with falls N=4,273 Profile (%)
11111 (19.2)	**11111 (57.5)**	**11111 (14.9)**	21121 (13.8)	21221 (10.5)	21221 (9.4)
21121 (10.4)	11121 (11.6)	21121 (10.7)	21221 (12.9)	21121 (9.5)	21121 (9.1)
21221 (9.1)	21111 (5.2)	21221 (9.7)	11121 (7.5)	**11111 (7.3)**	**11111 (6.8)**
11121 (8.8)	21121 (5.1)	11121 (7.6)	21222 (6.5)	21222 (6.6)	21222 (6.8)
21111 (5.2)	21221 (2.6)	21111 (5.7)	**11111 (6.3)**	11121 (5.6)	11121 (6.0)
21222 (4.5)	11112 (2.3)	21222 (5.0)	22221 (4.3)	22221 (4.7)	22222 (4.5)
22221 (3.1)	11122 (1.7)	22221 (3.8)	21111 (3.9)	21111 (4.2)	21111 (3.9)
21122 (2.8)	21211 (1.4)	21122 (3.1)	21122 (3.8)	22222 (3.9)	21122 (3.7)
21211 (2.4)	11211 (1.4)	21211 (3.0)	22222 (2.9)	21122 (3.7)	22221 (3.7)
22222 (2.1)	11221 (1.1)	22222 (2.5)	21211 (1.9)	21211 (2.5)	21211 (2.1)

[a] Digits one through five represent respectively: Mobility; Self-care; Usual activities; Pain/Discomfort; Anxiety/Depression; '1' indicates 'no problems', '2' indicates 'some problems', and '3' indicates 'extreme problems' with domain.
[b] Reported having none of the recorded 17 chronic conditions.
[c] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

The EQ-5D mobility and Katz walking items were moderately correlated in all subgroups (Table 4). The EQ-5D self-care item and Katz bathing and dressing items were very strongly correlated in the 'healthy' subgroup and strongly correlated in the geriatric conditions subgroups. A strong correlation was observed between the EQ-5D usual activities item and the Katz IADL summary score in all subgroups except for the dizziness with falls subgroup. In the

'healthy' subgroup, a moderate correlation was observed between the EQ-5D anxiety/depression item and psychological well-being (as measured by the Rand-36 mental health subscale score) and a weak correlation was observed between the EQ-5D summary score and a general quality of life score (as measured by Cantril's Self-anchoring Ladder). This differed from the geriatric condition subgroups which demonstrated strong and moderate correlations for these respective measures.

Table 4 Random effects meta-correlation coefficient (Spearman's rho) between EQ-5D select dimensions/summary score and related survey items

	Overall sample N=25,637 ϱ (95% CI)	**Healthy[a] N=2,275** ϱ (95% CI)	**Hearing issues N=9,762** ϱ (95% CI)	**Joint damage[b] N=11,903** ϱ (95% CI)	**Urinary incontinence N=5,932** ϱ (95% CI)	**Dizziness with falls N=4,273** ϱ (95% CI)
Mobility and Katz mobility item 'Walking'	0.39 (0.35, 0.43)	0.46 (0.39, 0.53)	0.38 (0.33, 0.42)	0.33 (0.29, 0.37)	0.35 (0.31, 0.39)	0.38 (0.31, 0.44)
Self-care and Katz ADL item 'Bathing'	0.68 (0.65, 0.71)	0.83 (0.73, 0.89)	0.67 (0.63, 0.70)	0.64 (0.61, 0.68)	0.69 (0.64, 0.73)	0.68 (0.64, 0.71)
Self-care and Katz ADL item 'Dressing'	0.61 (0.57, 0.64)	0.80 (0.69, 0.88)	0.59 (0.55, 0.62)	0.57 (0.53, 0.61)	0.61 (0.57, 0.64)	0.61 (0.57, 0.65)
Usual Activities and IADL summary score	0.51 (0.47, 0.55)	0.55 (0.41, 0.66)	0.51 (0.47, 0.55)	0.51 (0.46, 0.55)	0.51 (0.46, 0.55)	0.47 (0.43, 0.52)
Anxiety / Depression and Psychological wellbeing[c]	-0.52 (-0.55, -0.49)	-0.36 (-0.42, -0.29)	-0.52 (-0.56, -0.48)	-0.53 (-0.57, -0.50)	-0.55 (-0.59, -0.52)	-0.58 (-0.62, -0.54)
EQ-5D summary score and General quality of life score[d]	0.40 (0.35, 0.45)	0.28 (0.19, 0.37)	0.38 (0.34, 0.43)	0.38 (0.32, 0.43)	0.39 (0.34, 0.43)	0.40 (0.34, 0.47)

Abbreviations: IADL, Instrumental Activities of Daily Living; ADL, Activities of Daily Living

[a] Reported having none of the 17 chronic conditions recorded in The Older Persons and Informal Caregivers Survey.

[b] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

[c] Based on the Rand mental health subscale.

[d] Based on Cantril's Self Anchoring Ladder.

In the overall study population, the average EQ-5D summary score was higher in respondents who were younger, male, married, more highly educated, and residing with others (Table 5). A clear gradient was observed by co-morbidity status, with fewer co-morbidities resulting in improved HRQoL scores. When adjusted for age and sex, associations were broadly similar to the unadjusted model. Similar patterns were observed across subgroups in the unadjusted and adjusted models (Supplemental Tables 1 and 2).

Table 5 Unadjusted and adjusted EQ-5D summary scores for the overall population (N=25,637)

	Unadjusted	Adjusted for age and sex[b]
Age		
Mean age[a]	**0.69 (0.66, 0.73)**	--
Per additional year	-0.003 (-0.003, -0.002)	--
Sex		
Men (reference)	**0.73 (0.70, 0.77)**	--
Women	-0.06 (-0.07, -0.06)	--
Marital status		
Married/cohabiting (reference)	**0.71 (0.68, 0.74)**	**0.73 (0.70, 0.77)**
Widowed	-0.03 (-0.04, -0.03)	-0.0001 (-0.01, 0.01)
Single/divorced	-0.03 (-0.04, -0.02)	-0.01 (-0.02, -0.002)
Educational level		
Primary	-0.03 (-0.04, -0.02)	-0.02 (-0.03, -0.02)
Secondary (reference)	**0.70 (0.67, 0.74)**	**0.73 (0.70, 0.76)**
College/Some college	0.03 (0.02, 0.04)	0.02 (0.01, 0.03)
Living arrangements		
Lives alone	-0.03 (-0.03, -0.02)	-0.002 (-0.01, 0.005)
Lives with others (reference)	**0.71 (0.67, 0.74)**	**0.73 (0.70, 0.76)**
Hearing issues		
Yes (reference)	**0.67 (0.65, 0.71)**	**0.72 (0.68, 0.75)**
No	0.03 (0.02, 0.03)	0.02 (0.02, 0.03)
Joint damage[c]		
Yes (reference)	**0.63 (0.60, 0.67)**	**0.66 (0.63, 0.69)**
No	0.12 (0.11, 0.12)	0.11 (0.10, 0.11)

Table 5 Continued

	Unadjusted	Adjusted for age and sex[b]
Urinary incontinence		
Yes (reference)	**0.62 (0.58, 0.65)**	**0.66 (0.62, 0.69)**
No	0.10 (0.10, 0.11)	0.09 (0.08, 0.10)
Dizziness with falls		
Yes (reference)	**0.59 (0.56, 0.62)**	**0.63 (0.60, 0.66)**
No	0.13 (0.12, 0.13)	0.12 (0.11, 0.13)
Other co-morbidities[d]		
One co-morbidity	0.20 (0.19, 0.21)	0.19 (0.18, 0.20)
Two co-morbidities	0.15 (0.14, 0.15)	0.15 (0.14, 0.15)
Three co-morbidities	0.10 (0.09, 0.11)	0.10 (0.09, 0.11)
Four or more co-morbidities (reference)	**0.59 (0.56, 0.62)**	**0.63 (0.60, 0.66)**

[a] Mean age was 78 years.
[b] References are based on men and the centered mean age for healthy/morbidity subgroups.
[c] Defined as arthrosis or degenerative arthritis of the hips and/or knees.
[d] Included self-reported diabetes, cerebrovascular events, heart failure, cancer, airway disease, osteoporosis, bone fractures, depression, anxiety, dementia, or vision problems.

Discussion

This study examined the measurement of properties of the EQ-5D across four major geriatric subgroups and supports its validity in the context of TOPICS-MDS. Although the overall study population suggests a ceiling effect in the EQ-5D, this was driven, in part, by a 'healthy' subpopulation imbedded with the database. When consideration was given to major geriatric conditions (in essence, as different index conditions) the EQ-5D was found to confer adequate discriminative ability. Reassuringly, the majority of hypotheses regarding correlations between survey items and sociodemographic characteristics were supported, suggesting that construct validity was maintained. Strong correlations were observed between the EQ-5D self-care item and the Katz Index items for bathing and dressing as well as between the EQ-5D usual activities item and the Katz IADL summary score. Strong correlations were also observed between the EQ-5D anxiety/depression item and the Rand-36 mental

health sub-scale summary score. Moderate correlations were observed between the EQ-5D summary and general quality of life scores for the overall study population as well as across morbidity subgroups. Furthermore, EQ-5D scores were lower in respondents who were older, female, widowed/single, lower educated, and living alone. Increasing co-morbidity had a clear negative effect on EQ-5D scores.

However, there were several notable observations regarding the instrument's performance. Previous research has reported the weak performance of the EQ-5D in older persons with hearing issues[33] and has suggested that the Health Utility Index Mark III (HUI3) possessed better discriminatory ability for HRQoL reporting in this subgroup.[33,34] Although a borderline ceiling effect was observed in this study, arguably, the EQ-5D still provided an adequate measure of HRQoL in this subgroup. To further test the robustness of the EQ-5D in older persons with hearing issues, future research would ideally assess the responsiveness of the instrument in the context of TOPICS-MDS.

Since the presence of ceiling effects is a well-recognized limitation of the EQ-5D, it was unsurprising that this instrument lacked discriminative ability for 'healthy' older persons. However, it is important to emphasize that the database contains relatively few respondents who fall into this category (<10%). Most older persons reported multiple conditions, and for these respondents, the EQ-5D provides a suitable assessment to discriminate HRQoL.

There were several cases where convergent validity deviated from *a priori* hypotheses. In all subgroups, only moderate correlations were observed between the EQ-5D mobility item and the Katz walking item. In the dizziness with falls subgroup, there was only a moderate correlation between the EQ-5D usual activities item and the IADL summary score. In the 'healthy' subgroup, a weak correlation was observed between the EQ-5D summary score and the general quality of life score. It is uncertain why these findings arose. Speculatively, the Katz walking item may lack sufficient detail for older persons with multimorbidity. Whereas some may argue that that the EQ-5D usual activities domain and the IADL summary score represent conceptually different constructs, it remains unclear why the strength of the correlation would only be affected in the dizziness in falls subgroup. In the healthy subgroup, ceiling effects likely attenuated the correlation between the EQ-5D summary score and the general quality of life score (i.e. Cantril's Self Anchoring Ladder).

Several limitations must be noted. Firstly, data on geriatric conditions were self-reported posing the risk of reporting bias. Underreporting of hearing loss[35] and urinary incontinence[36] in older persons is a well-known phenomenon and may be partly attributed to social embarrassment. Furthermore, there are knowledge gaps in older persons' understanding of urinary incontinence[37];

involuntary loss of urine during physical exertion or laughing is not always recognized as a form of incontinence. For these reasons, it is likely that these two conditions were underreported in our study. Whereas older persons may not always report dizziness to their health care providers[38], this does not necessarily indicate that this item would be underreported in this survey. Similarly, joint damage is less prone to reporting bias.[39] Nonetheless, in the absence of complementary clinical data, the magnitude of reporting bias for each of these conditions could not be discerned.

Secondly, data on geriatric conditions were only reported in the broadest sense, i.e. presence or absence of the condition. This lack of specificity precluded the examination of known-group validity between respondents with differing severity levels. However, this limitation is not distinct to TOPICS-MDS; many general health surveys are not inherently designed to extract detailed information on specific conditions. Similar to these surveys, TOPICS-MDS was designed to provide a more global perspective of health and wellbeing status. In this regard, it is also important to emphasize that TOPICS-MDS only captures information on 17 chronic conditions common in older populations. A substantial proportion of older persons classified as 'healthy' did not report optimal health-related quality of life, and thus, the 'healthy' subgroup may have had acute or chronic conditions not captured in this survey. Nonetheless, this subgroup did report higher HRQoL, psychological wellbeing, and general quality of life, suggesting a healthier segment of the population.

Moreover, heterogeneity within TOPICS-MDS is a concern since it is a pooled dataset comprised of research projects with different protocols and sampling frameworks (e.g. samples taken from primary care centers, hospital settings and the general population). To address this issue, meta-analytic techniques were applied. Random effects by research project were included in both the correlation analyses and linear mixed models.

Lastly, in the primary analysis, morbidity subgroups were not mutually exclusive, potentially biasing the interpretation of findings. However, this limitation is not distinct to this dataset, but is rather a widespread issue when conducting morbidity research in older persons. In contemporary medicine, older persons rarely present with a single chronic condition but rather a range of conditions.[40] To address this issue in this study, individual geriatric conditions were regressed on EQ-5D summary scores in the different subgroups to examine the impact of individual conditions as well as co-morbidity on the outcome.

In the arena of quality of life research, this study is highly relevant as it performed a thorough analysis of measurement properties of the EQ-5D across four major geriatric conditions in a large group of older persons and supported

the discriminative ability of the EQ-5D. The unique infrastructure of TOPICS-MDS allowed for the pooled analysis of individual patient data from 32 research projects, which in turn granted the opportunity to explore subgroup analyses. Such findings are not only pertinent to users of TOPICS-MDS but also to the broader research community interested in accurate wellbeing measures for use in older populations. Furthermore, this study underlines that TOPICS-MDS, as a data sharing initiative, collects key variables for assessing quality of life and wellbeing in older persons.

Conclusion

This study supported the construct validity of the EQ-5D in the overall TOPICS-MDS study population as well as across subgroups of older persons presenting with four major geriatric conditions: hearing issues, joint damage, urinary incontinence, and dizziness with falls. Relative to the other three conditions, the risk of ceiling effects was higher for persons with hearing issues. For older persons who are generally healthy, i.e. reporting few to no chronic conditions, the EQ-5D confers poor discriminative ability. Although the pooled dataset for TOPICS-MDS may initially suggest poor discriminative ability for the EQ-5D, such is not the case when a healthy subgroup is distinguished from subgroups presenting with major geriatric conditions.

References

1. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *Lancet*. Nov 6 2014.
2. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics*. Jan 2000;17(1):13-35.
3. Tamburini M. Health-related quality of life measures in cancer. *Ann Oncol*. 2001;12 Suppl 3:S7-10.
4. EuroQoL Research Foundation. What is EQ-5D? Available at: http://www.euroqol.org/. Accessibility verified October 16, 2015.
5. Bharmal M, Thomas J, 3rd. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health*. Jul-Aug 2006;9(4):262-271.
6. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. *Value Health*. Mar-Apr 2011;14(2):390-399.
7. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. Sep 2004;13(9):873-884.
8. Tordrup D, Mossman J, Kanavos P. Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented? *Int J Technol Assess Health Care*. Jan 2014;30(1):10-19.
9. Moller AH, Erntoft S, Vinding GR, Jemec GB. A systematic literature review to compare quality of life in psoriasis with other chronic diseases using EQ-5D-derived utility values. *Patient Relat Outcome Meas*. 2015;6:167-177.
10. Hunger M, Thorand B, Schunk M, et al. Multimorbidity and health-related quality of life in the older population: results from the German KORA-age study. *Health Qual Life Outcomes*. 2011;9:53.
11. Davis S, Wailoo A. A review of the psychometric performance of the EQ-5D in people with urinary incontinence. *Health Qual Life Outcomes*. 2013;11:20.
12. Grutters JP, Joore MA, van der Horst F, Verschuure H, Dreschler WA, Anteunis LJ. Choosing between measures: comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. *Qual Life Res*. Oct 2007;16(8):1439-1449.
13. Streiner DL, Norman GR. *Health Measure Scales: a pratical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008:167-210; 247-276.
14. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. *J Am Geriatr Soc*. May 2007;55(5):780-791.
15. Lutomski JE, Baars MA, Schalk BW, et al. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PloS One*. 2013;8(12):e81673.
16. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37:53-72.
17. Lamers LM, Stalmeier PFM, McDonnell J, Krabbe PFM, van Busschbach JJ. [Measuring the quality of life in cost-utility analyses: the Dutch EQ-5D tariff]. *Ned Tijdschr Geneeskd*. 2005;149(28):1574-1578.
18. Monitor Gezondheid. Lokale en nationale monitor gezondheid (*in Dutch only*). Available at: https://www.monitorgezondheid.nl/gezondheidindicatoren.aspx. Accessibility verified September 13, 2015.
19. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. *JAMA*. Sep 21 1963;185:914-919.
20. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc*. Jun 1992;40(6):585-588.
21. van der Zee K, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding (*available in Dutch only*) Available at: http://www.rug.nl/research/share/research/tools/assessment-tools/rand36. Accessibility verified February 26, 2013.
22. Cantril H. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press; 1965.
23. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof*. Mar 2002;25(1):76-97.

24. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* Jan 2007;60(1):34-42.
25. Schwarzer G. 2013; *meta*: Meta-analysis with R, http://cran.r-project.org/web/packages/meta/meta.pdf, Version = 3.1-2
26. Hopkins WG. A new view of statistics: Effect magnitudes. 2002; Available at: http://www.sportsci.org/resource/stats/effectmag.html. Accessibility verified October 16, 2015.
27. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Qual Life Res.* Jun 1993;2(3):169-180.
28. Han KT, Park EC, Kim JH, Kim SJ, Park S. Is marital status associated with quality of life? *Health Qual Life Outcomes.* 2014;12:109.
29. Cherepanov D, Palta M, Fryback DG, Robert SA. Gender differences in health-related quality-of-life are partly explained by sociodemographic and socioeconomic variation between adult men and women in the US: evidence from four US nationally representative data sets. *Qual Life Res.* Oct 2010;19(8):1115-1124.
30. Hawton A, Green C, Dickens AP, et al. The impact of social isolation on the health status and health-related quality of life of older people. *Qual Life Res.* Feb 2011;20(1):57-67.
31. Agborsangaya CB, Lau D, Lahtinen M, Cooke T, Johnson JA. Health-related quality of life and healthcare utilization in multimorbidity: results of a cross-sectional survey. *Qual Life Res.* May 2013;22(4):791-799.
32. Mujica-Mota RE, Roberts M, Abel G, et al. Common patterns of morbidity and multi-morbidity and their impact on health-related quality of life: evidence from a national survey. *Qual Life Res.* Apr 2015;24(4):909-918.
33. Yang Y, Longworth L, Brazier J. An assessment of validity and responsiveness of generic measures of health-related quality of life in hearing impairment. *Qual Life Res.* Dec 2013;22(10):2813-2828.
34. Kuthubutheen J, Mittmann N, Amoodi H, Qian W, Chen JM. The effect of different utility measures on the cost-effectiveness of bilateral cochlear implantation. *Laryngoscope.* Feb 2015;125(2):442-447.
35. Kiely KM, Gopinath B, Mitchell P, Browning CJ, Anstey KJ. Evaluating a dichotomized measure of self-reported hearing loss against gold standard audiometry: prevalence estimates and age bias in a pooled national data set. *J Aging Health.* Apr 2012;24(3):439-458.
36. Shah D, Badlani G. Treatment of overactive bladder and incontinence in the elderly. *Rev Urol.* 2002;4 Suppl 4:S38-43.
37. Branch LG, Walker LA, Wetle TT, DuBeau CE, Resnick NM. Urinary incontinence knowledge among community-dwelling people 65 years of age and older. *J Am Geriatr Soc.* Dec 1994;42(12):1257-1262.
38. Bird JC, Beynon GJ, Prevost AT, Baguley DM. An analysis of referral patterns for dizziness in the primary care setting. *Br J Gen Pract.* Dec 1998;48(437):1828-1832.
39. Baumeister H, Kriston L, Bengel J, Harter M. High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. *J Clin Epidemiol.* May 2010;63(5):558-565.
40. Banerjee S. Multimorbidity-older adults need health care that can count past one. *Lancet.* Nov 6 2014.

Supplemental Table 1a Unadjusted EQ-5D summary scores by geriatric condition

	Healthy[a]	Hearing issues	Joint damage[b]	Urinary incontinence	Dizziness with falls
Age					
Mean age[c]	**0.84 (0.81, 0.88)**	**0.67 (0.64, 0.71)**	**0.64 (0.61, 0.67)**	**0.61 (0.58, 0.65)**	**0.59 (0.55, 0.62)**
Per additional year	-0.002 (-0.003, -0.001)	-0.003 (-0.004, -0.003)	-0.003 (-0.003, -0.002)	-0.002 (-0.003, -0.001)	-0.0003 (-0.002, 0.001)
Sex					
Men (reference)	**0.85 (0.82, 0.89)**	**0.71 (0.68, 0.75)**	**0.68 (0.64, 0.71)**	**0.64 (0.61, 0.68)**	**0.61 (0.57, 0.65)**
Women	-0.02 (-0.03, 0.01)	-0.07 (-0.08, -0.06)	-0.05 (-0.06, -0.04)	-0.03 (-0.05, -0.01)	-0.03 (-0.05, -0.02)
Marital status					
Married / cohabiting (reference)	**0.85 (0.81, 0.88)**	**0.69 (0.65, 0.72)**	**0.65 (0.62, 0.69)**	**0.63 (0.60, 0.67)**	**0.59 (0.55, 0.63)**
Widowed	-0.02 (-0.03, -0.003)	-0.03 (-0.04, 0.02)	-0.02 (-0.03, -0.01)	-0.02 (-0.03, -0.004)	-0.01 (-0.03, 0.01)
Single / divorced	-0.004 (-0.03, 0.02)	-0.03 (-0.05, -0.01)	-0.02 (-0.03, -0.003)	-0.009 (-0.03, 0.01)	-0.01 (-0.04, 0.02)
Educational level					
Primary	-0.005 (-0.02, 0.01)	-0.03 (-0.04, -0.02)	-0.04 (-0.05, -0.03)	-0.03 (-0.05, -0.01)	-0.02 (-0.04, -0.003)
Secondary (reference)	**0.84 (0.81, 0.88)**	**0.68 (0.65, 0.71)**	**0.65 (0.62, 0.69)**	**0.63 (0.59, 0.67)**	**0.59 (0.55, 0.63)**
College / Some college	0.002 (-0.01, 0.02)	0.03 (0.02, 0.05)	0.03 (0.01, 0.04)	0.04 (0.02, 0.06)	0.03 (0.01, 0.06)
Living arrangements					
Lives alone	-0.02 (-0.03, -0.01)	-0.03 (-0.04, -0.02)	-0.01 (-0.02, -0.004)	-0.01 (-0.03, 0.001)	-0.01 (-0.03, 0.01)
Lives with others (reference)	**0.85 (0.81, 0.89)**	**0.68 (0.65, 0.72)**	**0.65 (0.62, 0.68)**	**0.63 (0.59, 0.66)**	**0.59 (0.55, 0.63)**

[a] Reported having none of the 17 chronic conditions recorded in The Older Persons and Informal Caregivers Survey.
[b] Defined as arthrosis or degenerative arthritis of the hips and/or knees.
[c] Mean ages were 75, 80, 78, 79, and 79 for healthy, hearing issues, joint damage, urinary incontinence and dizziness with falls subgroups respectively.

Supplemental Table 1b Unadjusted EQ-5D summary scores by geriatric condition

	Hearing issues	Joint damage[a]	Urinary incontinence	Dizziness with falls
Hearing issues				
Yes (reference)	--	**0.63 (0.60, 0.67)**	**0.62 (0.58, 0.65)**	**0.58 (0.54, 0.61)**
No	--	0.02 (0.01, 0.03)	0.01 (-0.003, 0.03)	0.02 (0.001, 0.04)
Joint damage				
Yes (reference)	**0.62 (0.58, 0.65)**	--	**0.58 (0.54, 0.62)**	0.55 (0.51, 0.58)
No	0.12 (0.11, 0.13)	--	0.10 (0.09, 0.11)	0.09 (0.07, 0.11)
Urinary incontinence				
Yes (reference)	**0.61 (0.57, 0.64)**	**0.58 (0.55, 0.61)**	--	**0.53 (0.49, 0.57)**
No	0.10 (0.09, 0.11)	0.08 (0.07, 0.09)	--	0.09 (0.08, 0.11)
Dizziness with falls				
Yes (reference)	**0.57 (0.54, 0.61)**	**0.55 (0.53, 0.59)**	**0.53 (0.50, 0.57)**	--
No	0.13 (0.12, 0.14)	0.10 (0.09, 0.12)	0.12 (0.10, 0.13)	--
Other co-morbidities[b]				
0 co-morbidity	0.19 (0.18, 0.21)	0.16 (0.15, 0.18)	0.19 (0.16, 0.21)	0.23 (0.20, 0.26)
1 co-morbidities	0.15 (0.14, 0.16)	0.13 (0.12, 0.14)	0.15 (0.13, 0.17)	0.16 (0.14, 0.18)
2 co-morbidities	0.10 (0.09, 0.12)	0.09 (0.08, 0.11)	0.11 (0.09, 0.13)	0.13 (0.11, 0.15)
3+ co-morbidities (reference)	**0.58 (0.55, 0.61)**	**0.56 (0.53, 0.59)**	**0.53 (0.50, 0.56)**	**0.50 (0.46, 0.53)**

[a] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

[b] Included self-reported diabetes, cerebrovascular events, heart failure, cancer, airway disease, osteoporosis, bone fractures, depression, anxiety, dementia, or vision problems.

Supplemental Table 2a Age and sex adjusted[a] EQ-5D summary scores by sociodemographic characteristics

	Healthy[b]	Hearing issues	Joint damage[c]	Urinary incontinence	Dizziness with falls
Marital status					
Married/ cohabiting (reference)	**0.85 (0.82, 0.89)**	**0.71 (0.67, 0.74)**	**0.67 (0.64, 0.70)**	**0.64 (0.61, 0.68)**	**0.61 (0.57, 0.65)**
Widowed	-0.01 (-0.02, 0.01)	0.01 (-0.01, 0.02)	0.01 (-0.01, 0.02)	-0.002 (-0.02, 0.01)	0.003 (-0.02, 0.02)
Single/divorced	-0.001 (-0.02, 0.02)	-0.01 (-0.03, 0.01)	-0.01 (-0.02, 0.01)	-0.002 (-0.03, 0.02)	-0.004 (-0.03, 0.02)
Educational level					
Primary	0.00002 (-0.02, 0.02)	-0.02 (-0.03, -0.01)	-0.03 (-0.04, -0.02)	-0.03 (-0.04, -0.01)	-0.02 (-0.04, -0.001)
Secondary (reference)	**0.85 (0.82, 0.89)**	**0.71 (0.67, 0.74)**	**0.68 (0.64, 0.71)**	**0.64 (0.60, 0.68)**	**0.61 (0.57, 0.65)**
College/ Some college	-0.002 (-0.02, 0.01)	0.03 (0.01, 0.04)	0.02 (0.01, 0.03)	0.04 (0.02, 0.06)	0.03 (0.005, 0.05)
Living arrangements					
Lives alone	-0.01 (-0.02, 0.003)	0.01 (-0.002, 0.02)	0.01 (-0.002, 0.02)	-0.0004 (-0.02, 0.01)	0.003 (-0.02, 0.02)
Lives with others (reference)	**0.86 (0.82, 0.89)**	**0.70 (0.67, 0.74)**	**0.67 (0.64, 0.70)**	**0.64 (0.60, 0.68)**	**0.61 (0.57, 0.65)**

[a] References are based on men and the centered mean age for healthy/morbidity subgroups. Average ages were 75, 80, 78, 79, and 79 for healthy, hearing issues, joint damage, urinary incontinence and dizziness with falls subgroups respectively.

[b] Reported having none of the 17 chronic conditions recorded in The Older Persons and Informal Caregivers Survey.

[c] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

Supplemental Table 2b Age and sex adjusted[a] EQ-5D summary scores by sociodemographic characteristics

	Hearing issues	Joint damage[b]	Urinary incontinence	Dizziness with falls
Hearing issues				
Yes (reference)	--	**0.66 (0.63, 0.70)**	**0.64 (0.60, 0.68)**	**0.60 (0.56, 0.64)**
No	--	0.02 (0.01, 0.03)	0.01 (-0.01, 0.02)	0.02 (0.003, 0.04)
Joint damage				
Yes (reference)	**0.64 (0.61, 0.68)**	--	**0.59 (0.55, 0.62)**	**0.56 (0.52, 0.60)**
No	0.10 (0.09, 0.11)	--	0.10 (0.08, 0.11)	0.09 (0.07, 0.11)
Urinary incontinence				
Yes (reference)	**0.64 (0.61, 0.68)**	**0.61 (0.58, 0.65)**	--	**0.54 (0.50, 0.58)**
No	0.08 (0.07, 0.09)	0.07 (0.06, 0.09)	--	0.09 (0.07, 0.11)
Dizziness with falls				
Yes (reference)	**0.61 (0.57, 0.64)**	**0.59 (0.56, 0.62)**	**0.55 (0.52, 0.59)**	--
No	0.12 (0.11, 0.13)	0.10 (0.09, 11)	0.12 (0.10, 0.13)	--
Other co-morbidities[c]				
0 co-morbidities	0.19 (0.17, 0.20)	0.16 (0.15, 0.18)	0.19 (0.17, 0.21)	0.23 (0.20, 0.26)
1 co-morbidity	0.15 (0.13, 0.16)	0.13 (0.12, 0.14)	0.15 (0.14, 0.17)	0.16 (0.14, 0.18)
2 co-morbidities	0.10 (0.09, 0.11)	0.09 (0.08, 0.11)	0.11 (0.10, 0.13)	0.13 (0.11, 0.15)
3+ co-morbidities (reference)	**0.61 (0.58, 0.64)**	**0.59 (0.56, 0.62)**	**0.56 (0.53, 0.60)**	**0.52 (0.48, 0.55)**

[a] References are based on men and the centered mean age for healthy/morbidity subgroups. Average ages were 75, 80, 78, 79, and 79 for healthy, hearing issues, joint damage, urinary incontinence and dizziness with falls subgroups respectively.

[b] Defined as arthrosis or degenerative arthritis of the hips and/or knees.

[c] Included self-reported diabetes, cerebrovascular events, heart failure, cancer, airway disease, osteoporosis, bone fractures, depression, anxiety, dementia, or vision problems.

3.4

The impact of study design on reported care-related quality of life

Published as:

Lutomski JE, van Exel NJ, Kempen GI, Moll van Charante EP, den Elzen WP, Jansen AP, Krabbe PF, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis RJ on behalf of TOPICS-MDS Consortium. Validation of the Care-Related Quality of Life Instrument in different study settings: findings from The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS). *Quality of Life Research*. 2015 May;24(5):1281-93.

Abstract

Purpose: Validity is a contextual aspect of a scale which may be differ across sample populations and study protocols. The objective of our study was to validate the Care-Related Quality of Life Instrument (CarerQol) across two different study design features, sampling framework (general population versus different care settings) and survey mode (interview versus written questionnaire).
Methods: Data were extracted from The Older Persons and Informal Caregivers Minimum DataSet (TOPICS-MDS, www.topics-mds.eu), a pooled open access dataset with information on >3,000 informal caregivers throughout the Netherlands. Meta-correlations and linear mixed models between the CarerQol's seven dimensions (CarerQol-7D) and caregiver's level of happiness (CarerQol-VAS) and self-rated burden (SRB) were performed.
Results: The CarerQol-7D dimensions were correlated to the CarerQol-VAS and SRB in the pooled dataset and the subgroups. The strength of correlations between CarerQol-7D dimensions and SRB was weaker among caregivers who were interviewed versus those who completed a written questionnaire. The directionality of associations between the CarerQol-VAS, SRB and the CarerQol-7D dimensions in the multivariate model supported the construct validity of the CarerQoL in the pooled population. Significant interaction terms were observed in several dimensions of the CarerQoL-7D across sampling frame and survey mode, suggesting meaningful differences in reporting levels.
Conclusions: Although good scientific practice emphasizes the importance of re-evaluating instrument properties in individual research studies, our findings support the validity and applicability of the CarerQol instrument in a variety of settings. Due to minor differential reporting, pooling CarerQol data collected using mixed administration modes should be interpreted with caution; for TOPICS-MDS, meta-analytic techniques may be warranted.

Introduction

Informal care for older persons has long served as an integral part of the health care system. However, increased longevity coupled with decreased fertility has widened the ratio between very old persons requiring care and middle-aged persons who have historically provided informal care.[1] These demographic trends are further compounded by wider geographical dispersion of family members[2], the increasing number of women joining the labor market[2] and the increasing prevalence of minor functional impairment and chronic (co-) morbidity among older persons.[3-5] The combination of these population dynamics leads to fewer persons available to provide support for older persons with more complex care needs as well as increased physical and mental stress on caregivers.[6-9] As government policies shift towards informal care to offset increasing health care costs, understanding care-related burden has become increasingly important in public health and health policy arenas.[10]

With this backdrop, the Dutch National Care for the Elderly Programme funded the development and implementation of The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS; www.topics-mds.eu) to guide future policy decision making and health interventions for older persons and informal caregivers in the Netherlands.[11] TOPICS-MDS is a large, open-access, pooled dataset with information from various research projects across the country. Yet, challenges in the analysis of TOPICS-MDS remain. Although the same set of core questions were answered by all caregivers, heterogeneity between study protocols may result in differential measurement error which becomes obscured in the pooled dataset. For instance, TOPICS-MDS contains mixed-mode data, i.e. caregivers responded to the same core questions either in a face-to-face interview or in a written questionnaire. Measurement error may vary across these two survey modes due to differences in auditory versus visual processing or a preference to provide socially desirable answers when interviewed.[12] Such phenomena are well documented in other research settings.[13] Furthermore, variation in sampling frames (e.g. sampled from the general population versus primary centres) may elicit different response patterns.[14] Projecting the directionality and degree of measurement error induced by different study designs is difficult and may not be uniform across all variables.[12,15]

In light of these issues, the objective of our study was to validate the Care-Related Quality of Life Instrument (CarerQol) across two different study design features available in TOPICS-MDS, sampling framework (general population versus different care settings) and survey mode (interview versus written questionnaire). Notably, the CarerQol was the primary instrument used in

TOPICS-MDS to measure care-related quality of life. Although it has been validated in earlier research[16,17], these findings were based on small (<250 participants), specialized settings using a written questionnaire. Since construct validity is contextual[15], further validation work has been recommended.[17]

Methods

Data source

Data were derived from TOPICS-MDS (www.topics-mds.eu), which is a public access data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere.[11] Briefly, the Dutch National Care for the Elderly Programme (http://www.nationaalprogrammaouderenzorg.nl) was established in 2008 to promote proactive, integrated health care for older persons with complex care needs. As part of this national agenda, TOPICS-MDS was developed to collect uniform information from all research projects funded under the Programme. Thus, TOPICS-MDS consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardised protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to TOPICS-MDS, the database is dynamic in nature and thus regularly updated with new observations. Our present analysis uses the first version of the dataset available as of January 2013 and is based on 17 research projects with data available on 3,269 informal caregivers. TOPICS-MDS is a fully anonymised dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Centre Ethical Committee review reference number: CMO: 2012/120).

Informal caregivers

Informal caregivers were defined as individuals who provided long-term, unpaid care for another individual in their family, household or social network who has physical, mental or cognitive limitations. Sociodemographic information available on informal caregivers included sex, age, socioeconomic status[11], the caregiver's relationship with the care recipient, whether the caregiver resided with the care recipient, self-reported health and objective and subjective care-related burden.

Objective care-related burden refers to indicators for the size of the care task which can be independently measured, such as the frequency of care provision. In TOPICS-MDS, the average number of hours caregivers assisted the care recipient with household tasks, personal care, or practical support such as transport or financial/administrative duties over a one-week period prior to the assessment was recorded.

Subjective care-related burden refers to personal perceptions of care burden. Given that internalization of burden may vary between caregivers, caregivers in similar situations may report varying levels of burden, which in turn influences care-related quality of life. The CarerQol instrument was used to measure subjective care-related quality of life (Figure 1; original Dutch version available at www.topics-mds.eu). This instrument is comprised of two parts, the Carer-QoL-7D and the CarerQoL visual analogue scale (CarerQol-VAS). Modelled after the EuroQol-5D, the CarerQoL-7D was developed to measure seven dimensions of care-related burden which can then be used to derive a care-related quality of life summary score. These dimensions include care-related fulfilment; relational problems with the care recipient; mental health problems; physical health problems; problems completing daily activities; financial security; social support. Caregivers can describe their personal situation by responding whether they have 'no', 'some', 'a lot' of problems for each attribute.[16-18] To generate a single summary score for the CarerQol-7D, the two positive items (care-related fulfilment and social support) are reverse scored and a set of weights (also referred to as a 'tariff') are applied to each level of the seven dimensions. These weights were based on preferences derived from the Dutch general population.[19] The CarerQol-7D summary score represents a utility score for the care situation that ranges between zero (worst informal care situation) and 100 (best informal care situation).

The CarerQol-Visual Analogue Scale (CarerQol-VAS) represents an additional valuation component which asks caregivers to rate their general level of happiness using the statement "Please indicate how happy you are currently feeling". Responses are recorded on a horizontal scale ranging from '0' (completely unhappy) to '10' (completely happy). For the purposes of this analysis, scores were subsequently converted to a scale ranging from 0 to 100.

Two additional VAS were also assessed[20]: (1) self-rated care burden (SRB), and (2) transfer of care (Transfer). SRB is an independent measure of subjective care burden that provides an overall assessment of the negative and positive aspects of caregiving.[21] SRB was assessed using the following statement, "Draw an 'X' on the scale to indicate how hard you are finding it to care for the care recipient". Transfer describes a hypothetical situation in which care tasks are taken over by a person selected by the carer and the care recipient without

changing the living situation of the care recipient and free of charge.[22] Given this scenario, caregivers were asked to "Draw an 'X' on the scale to indicate how happy you would feel if this person would take over the care responsibilities from you". Unlike SRB, Transfer was not developed as a stand-alone measure but rather with the CarerQol-VAS to measure the concept of process utility (PU, described below). Both scales ranged from '0' to '10', and similarly to the CarerQol-VAS, were converted to scores ranging from 0 (not difficult at all, completely unhappy with transferring caregiving tasks) to 100 (far too difficult, completely happy with transferring caregiving tasks).

PU is a derived measure representing the value a carer attaches to the process of informal caregiving and is calculated by taking the difference in happiness between two situations: the current situation (measured by CarerQol-VAS) and a hypothetical situation in which care tasks are transferred (measured by Transfer).[22] A positive PU score indicates that the carer would be unhappy transferring care tasks, thus attaching a positive value to caregiving, whereas a negative PU score indicates the opposite. A PU score of zero indicates that that the carer is neutral towards caregiving.

Care recipient

Informal caregiver data were linked with data on the care recipient. To examine the relationship between caregiver quality of life and the severity of the care recipient's vulnerability, we constructed a 45-item frailty index [23] validated for use in TOPICS-MDS.[24] A frailty index is calculated by reviewing ≥30 health problems affecting different physiological systems; the total number of problems observed in an individual is then divided by the total number of problems reviewed to calculate a proportion. Frailty index scores ≥0.20 indicate a likely frail state.[23]

Statistical analysis

TOPICS-MDS is a pooled dataset of individual-level information; subsequent analyses were performed using a one-step individual patient data meta-analysis. This type of analysis, also referred to as a "mega-analysis", allows the simultaneous analysis of data while accounting for the clustering effects. Since TOPICS-MDS is comprised of study projects with different sampling frameworks and implementation procedures, we decided *a priori* to perform analyses in the pooled dataset as well as across two sub-groups: sampling framework (i.e. recruitment from the general population versus hospital versus primary care centres) and survey mode (i.e. caregivers who completed face-to-face interviews versus written questionnaires). Notably, individual study protocols dictated survey mode.

We first described the characteristics of the study sample. We then assessed the distributional properties of the CarerQoL-VAS, SRB and Transfer by examining mean scores, spread in variation (standard deviation), and the extent of floor or ceiling effects. Since there is no universally accepted threshold for identifying floor and ceiling effects, we considered these effects present if 15% of caregivers reported either the lowest or highest scores.25 We further examined the mean (standard deviation), range and floor and ceiling effects for PU and CarerQol-7D summary scores. Mean differences for the three VAS scores, PU and the CarerQol-7D summary score across subgroups were assessed using ANOVA.

For comparative purposes [17], we evaluated each of the seven dimensions of CarerQol-7D separately. We first calculated the distribution responses . To assess whether construct validity, i.e. the degree to which scores are consistent with hypotheses[26], of the CarerQoL-7D was upheld in the pooled dataset and predefined subgroups, we calculated random effects meta-correlations (Spearman's rho) between CarerQol-VAS, SRB and PU scores and the seven dimensions of the CarerQoL-7D with the *meta* package in R.[27] A random effects meta-correlation was performed to allow for heterogeneity between individual studies in the pooled dataset. Correlation coefficients were classified as trivial ($\leq$0.1), weak (0.1 to <0.3), moderate (0.3 to <0.5), strong (0.5 to <0.7), very strong ($\geq$0.7).[28] In line with previous research[17], we posited several hypotheses reading the directionality and magnitude of these associations (Supplemental Table 1). We presumed that construct validity was upheld if the strength of association and directionality of correlations coincided with most (~75%) of our hypotheses.[25]

Caregiver happiness and self-perceived burden have been shown to be associated with the CarerQoL-7D.[17] Thus, to further examine construct validity, we conducted linear mixed models to allow for clustering effects between study projects. Given the observed correlation between CarerQol-VAS and SRB, we conducted a multivariate repeated measures approach to simultaneously model both outcomes as a function of the seven dimensions of the CarerQoL-7D.[29] In the adjusted models, consideration was also given to caregiver's sex, age, socioeconomic status, self-reported health and the care recipient's sex, age and frailty status.

To demonstrate the potential impact on multivariate analysis, interaction terms were assessed to determine if associations differed significantly across sampling framework (general population versus different care settings) and survey mode (interview versus written questionnaire). Covariates of interest in these models only included the seven dimensions of the CarerQol-7D and their interaction terms between sampling framework or survey mode. All statistical analyses were carried out using SAS (Version 9.3; SAS Institute Inc., Care, NC, USA) and R (2013: Vienna, Austria).

CarerQol-7D

The following questions are about your situation as a caregiver. Place an "X" next to the word that best describes your care situation.

a. I gain…
☐ no
☐ some
☐ a lot
…(of) fulfilment with carrying out my care tasks.

b. I have…
☐ no
☐ some
☐ a lot
…(of) relational problems with the care receiver (for example, he/she is very demanding, he/she behaves differently, we have communication problems).

c. I have…
☐ no
☐ some
☐ a lot
…(of) problems with my own mental health (for example, stress, fear, gloominess, depression, concern about the future).

d. I have…
☐ no
☐ some
☐ a lot
…(of) problems with my own physical health (for example, being sick more often, tiredness, physical stress).

e. I have…
☐ no
☐ some
☐ a lot
…(of) problems combining my daily activities (for example, household activities, work, study, family activities and free time) with my care tasks.

f. I have…
☐ no
☐ some
☐ a lot
…(of) financial problems because of my care tasks.

g. I have…
☐ no
☐ some
☐ a lot
…(of) support with carrying out my care tasks when I need it (for example, from family, friends, neighbors, acquaintances).

CarerQol-VAS

Please draw an "X" on the scale below to indicate how happy you are feeling right now.

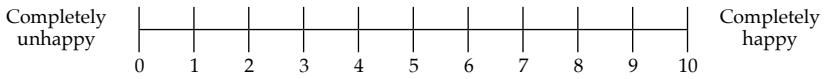


Figure 1 CarerQol instrument

Results

Sample characteristics

Data were pooled from 17 different research projects resulting in a dataset of 3,269 caregivers. The pooled dataset comprised of two large projects (>500 caregivers recruited), nine mid-sized projects (100 to <500 caregivers recruited) and six small projects (<100 caregivers recruited); a more detailed description of individual studies is available at www.topics-mds.eu. The majority of caregivers were recruited from primary care centres (59.3%; n=1,940), followed by the general population (23.0%; n=753) and hospitals (17.6%; n=576). Information on survey mode was available for two-third of caregivers (61.7%; n=2,195). Among caregivers with data on survey mode, most (86.3%; n=1,895) completed a written questionnaire. The majority of caregivers completing a written questionnaire were recruited from a primary care centre (59.5%; n=1,127).

Several differences in caregiver and care recipient characteristics were observed across sampling frames and survey mode (Table 1). Relative to those sampled from the general population or primary care centers, caregivers sampled from hospitals were more likely to be the care recipient's spouse and consequently reside with the care recipient. Care recipients sampled from hospitals were less likely to be female and on average reported a lower frailty index score. Caregivers who were interviewed were more likely to be older, the care recipient's spouse and reside with the care recipient; their care recipients were more likely to have at least one ADL limitation and to report higher frailty index scores.

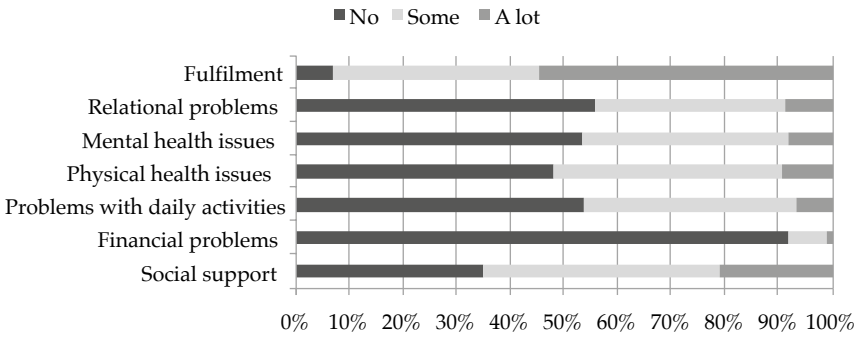


Figure 2 Distribution of CarerQol-7D dimension (N=3,269)

Table 1 Characteristics of informal caregivers by sampling frame and survey mode, The Older Persons and Informal Caregivers Minimum dataset, 2013

	Overall N=3,269	Sampling frame			Survey Mode		
		General population N=753	Hospital N=576	Primary care N=1,940	Interview N=300	Questionnaire N=1,895	Not indicated N=1,074
Caregiver							
Women	**70.1**	68.5	68.1	71.3	69.0	69.4	71.8
Age in years (mean, SD)	**63 (12)**	64 (12)	65 (12)	63 (13)	66 (12)	63 (12)	63 (12)
Relationship with care recipient							
Spouse/life partner	**43.9**	46.0	53.9	40.2	57.0	41.3	44.7
Child	**46.0**	43.2	38.5	49.2	36.0	47.0	47.0
Other	**10.1**	10.9	7.6	10.6	7.0	11.7	8.3
Residence with care recipient	**46.0**	43.9	54.4	44.3	60.1	44.5	44.6
Care hours/week (median, IQR)	**8 (16)**	9 (17)	11 (18)	8 (16)	12 (25)	9 (16)	7 (13)
Socioeconomic status quartile							
Low	**27.2**	19.8	35.4	27.6	19.6	32.3	20.3
Lower middle	**22.7**	17.8	22.4	24.7	25.3	21.2	24.6
Upper middle	**22.3**	21.3	14.5	24.9	31.8	21.7	20.6
High	**27.9**	41.2	27.7	22.8	23.3	24.8	34.5
Self-reported health							
Excellent/Very good	**24.2**	25.1	25.8	23.4	24.0	24.0	24.7
Good	**45.0**	45.4	44.2	45.1	46.3	45.4	44.0
Fair/Poor	**30.8**	29.5	30.0	31.5	29.7	30.6	31.4
Care recipient							
Women	**61.4**	61.6	55.4	63.1	52.3	63.6	60.2

Age in years (mean, SD)	**80 (7)**	80 (8)	78 (7)	80 (7)	80 (7)	80 (7)	80 (7)
Multimorbidity	**85.1**	81.3	77.9	87.8	91.3	86.9	80.8
≥1 ADL limitation	**89.3**	83.3	84.0	94.0	95.1	89.0	87.4
Frailty index score (mean, SD)	**0.33 (0.14)**	0.33 (0.16)	0.26 (0.14)	0.34 (0.13)	0.35 (0.14)	0.31 (0.13)	0.33 (0.16)

Abbreviations: SD, standard deviation; IQR, Interquartile range; ADL, Activities of Daily Living
Notes: Percentages are shown unless otherwise stated. Socioeconomic status was based on an index score derived from geospatial data on average income, employment type and educational level for the Dutch population. Multimorbidity was defined as the presence of >2 conditions based on a list of 17 morbidities. ADL limitations were identified using an extended version of the Katz Index of Independence Activities of Daily Living. The frailty index ranges from 0 to 1 with higher scores indicating higher levels of frailty.

Visual analogue scales and summary scores

Mean (SD) scores for the CarerQoL-VAS, SRB and Transfer were 70.1 (16.5), 43.0 (25.4) and 50.0 (26.2) respectively (Table 2). Based on a 15% threshold, no disproportionate clustering of responses was observed at the scale extremes to indicate floor or ceiling effects. However, for SRB, a notably higher proportion of caregivers sampled from hospitals reported no self-perceived burden (11.0%) relative to those sampled from the general population (2.3%) or primary care centers (3.3%). None of the caregivers who were interviewed reported the highest possible score for Transfer whereas 2.4% of caregivers who completed a written questionnaire did.

The distributional properties of PU and CarerQoL-7D summary scores were broadly similar across sampling frame and survey mode (Table 2). The mean (SD) PU score was 20.0 (31.3), indicating that on average caregivers attached a positive value to providing care. None of the caregivers scored the lowest or the highest possible scores for the CarerQol-7D (i.e. 0, 100 respectively). When examining lower and upper thresholds (i.e. summary scores $\leq$5 or $\geq$95), overall, no caregivers scored $\leq$5 whereas 6.9% scored $\geq$95.

Dimensions of the CarerQol

The vast majority of caregivers reported fulfilment in care-giving, and few reported financial difficulties due to their care-giving responsibilities (Figure 2). Reporting levels across the individual dimensions of the CarerQol-7D varied minimally by sampling frame and survey mode (Supplemental Figures 1 and 2); yet, there were several key differences. Caregivers sampled from primary care centres were less likely to report "no" relational issues compared to caregivers sampled from the general population or hospitals (50.2%, 63.7%, 64.9% respectively) and more likely to report "some" relational issues (40.1%, 28.6%, 27.9% respectively). Compared to those who completed a written questionnaire, caregivers who were interviewed reported higher levels of social support and relational problems though lower levels of fulfilment in caregiving.

Random effects meta-correlation

In the pooled dataset, a weaker negative correlation was observed between the CarerQol-VAS and SRB whereas a stronger positive correlation was observed between the CarerQol-VAS and PU (Table 3). There was a weak to moderate negative correlation between the CarerQol-VAS and five of the CarerQol-7D dimensions (relational problems, mental health issues, physical health issues, problems with daily activities and financial problems). SRB was negatively correlated with PU and fulfilment and positively correlated with the remaining CarerQol-7D dimensions. Similar patterns were observed for PU, though in

contrasting directionality. Observed correlations in the pooled dataset generally coincided with hypothesized magnitude and directionality (Table 3, Supplemental Table 1). The magnitudes of observed correlations differed modestly by sampling frame and survey mode. Specifically, correlations between SRB and the CarerQol-7D dimensions were weaker among caregivers who were sampled from primary care centres and who were interviewed (Supplemental Tables 2 and 3).

Multivariate analysis

The directionality of associations in the multivariate model between the CarerQol-VAS, SRB and the seven CarerQol dimensions supported the construct validity of the CarerQoL in the pooled population (Table 4). The positive dimensions of the CarerQol-7D (fulfilment, social support) were positively associated with the CarerQol-VAS whereas the negative dimensions (relational problems, mental health issues, physical health issues, problems with daily activities and financial problems) were negatively associated with the CarerQol-VAS; contrasting associations were observed for SRB. Similar associations persisted in the adjusted model. Poor self-rated health and frailty of the care recipient were negatively associated with the CarerQol-VAS and positively associated with increased SRB, further supporting the validity of the measures.

In the last analysis, the CarerQol-VAS and SRB were modelled as a function of the seven dimensions of the CarerQoL-7D and their interaction terms with sampling frame or survey mode; significant interactions terms were primarily observed for SRB (Tables 5 and 6). The patterns of association between the positive and negative dimensions of the CarerQol-7D and the CarerQol-VAS and SRB were similar to the unadjusted model. However, the interaction terms did impact the strength of these associations. Relative to caregivers sampled from primary care centres, there was a weaker association between relational problems and SRB in caregivers sampled from hospitals; in contrast, the association between mental health issues and SRB was stronger (Table 5). Moreover, a stronger association between problems with daily activities and SRB was observed for caregivers sampled from the general population. A significant interaction between survey mode and physical health was observed for both the CarerQol-VAS and SRB, underscoring meaningful differences in the reporting of this item in an interview versus a written questionnaire (Table 6). Relative to caregivers who completed a written questionnaire, a weaker association was found between physical health issues and the CarerQol-VAS and SRB in caregivers who were interviewed. Similarly, the association between problems with daily activities and SRB was weaker.

Table 2 Distribution of visual analogue scales, process utility and CarerQol-7D summary scores by sampling frame and survey mode

	Overall N=3,269	Sampling frame				Survey Mode			
		General population N=753	Hospital N=576	Primary care N=1,940	p-value	Interview N=300	Questionnaire N=1,895	Not indicated N=1,074	p-value
CarerQol-VAS									
Mean (SD)	**70.1 (16.5)**	71.3 (17.4)	68.0 (19.2)	70.3 (14.7)	0.07	68.7 (13.5)	69.7 (16.7)	72.3 (17.0)	0.11
Floor (%)	**0.3**	0	0.8	0.2		0	0.4	0	
Ceiling (%)	**3.6**	6.0	3.4	2.5		1.3	3.2	6.6	
SRB									
Mean (SD)	**43.0 (25.4)**	39.4 (24.5)	37.4 (26.2)	46.0 (25.1)	0.58	57.2 (25.7)	39.5 (24.9)	45.1 (24.5)	0.34
Floor (%)	**4.3**	2.3	11.0	3.3		4.0	6.5	0.7	
Ceiling (%)	**0.6**	1.0	0.4	0.5		0.3	0.4	1.0	
Transfer									
Mean (SD)	**50.0 (26.2)**	49.4 (25.8)	49.7 (27.5)	50.5 (26.0)	0.83	48.7 (26.5)	50.4 (26.2)	49.6 (26.1)	0.30
Floor (%)	**3.3**	1.8	5.9	3.2		4.4	4.1	0	
Ceiling (%)	**2.3**	2.6	2.4	2.0		0	2.4	3.1	
Process utility									
Mean (SD)	**20.0 (31.3)**	22.0 (30.6)	17.9 (33.1)	19.6 (30.9)	0.57	20.1 (31.1)	19.2 (31.5)	22.7 (30.5)	0.07
Range	**-90 to 100**	-90 to 100	-80 to 100	-90 to 100		-50 to 100	-90 to 100	-90 to 90	
Score <0	**20.9**	18.2	23.7	21.3		22.2	21.8	17.2	
Score =0	**17.5**	16.4	18.8	17.5		15.4	18.0	16.8	
Score>0	**61.6**	65.3	57.5	61.2		62.5	60.2	66.0	
CarerQol-7D									
Mean (SD)	**79.2 (14.7)**	81.0 (13.5)	79.3 (14.9)	78.5 (15.1)	0.98	78.1 (15.5)	80.5 (14.0)	77.3 (15.4)	0.95

Range	**14 to 98**	19 to 98	21 to 98	14 to 98	21 to 98	19 to 98	13 to 98
Summary score ≤5 (%)	**0**	0	0	0	0	0	0
Summary score ≥95 (%)	**6.9**	6.8	6.9	6.9	9.2	7.6	5.0

Abbreviations: SD, standard deviation
Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". Transfer refers to the statement, "Please indicate how happy you would feel if your caregiver responsibilities were taken over by someone else". CarerQol-VAS, SRB and Transfer are visual analogue scales ranging from 0 to 100. Process utility is a derived measure representing the value a carer attaches to the process of informal caregiving. Statistical significance between sampling frame and survey mode subgroups were tested using ANOVA.

Table 3 Random effects meta-correlation coefficient (Spearman's rho)

	CarerQol-VAS (95% CI)	**SRB (95% CI)**	**PU (95% CI)**
SRB	-0.25 (-0.40, -0.10)	--	--
PU	0.52 (0.49, 0.54)	-0.35 (-0.44, -0.24)	--
CarerQol domains			
Fulfilment	0.19 (0.15, 0.24)	-0.16 (-0.22, -0.09)	0.28 (0.22, 0.34)
Relational problems	-0.27 (-0.32, -0.22)	0.36 (0.27, 0.44)	-0.30 (-0.33, -0.27)
Mental health issues	-0.43 (-0.48, -0.38)	0.36 (0.26, 0.46)	-0.29 (-0.33, -0.25)
Physical health issues	-0.35 (-0.38, -0.31)	0.33 (0.21, 0.45)	-0.26 (-0.31, -0.21)
Problems with daily activities	-0.27 (-0.31, -0.24)	0.40 (0.27, 0.52)	-0.32 (-0.37, -0.28)
Financial problems	-0.17 (-0.21, -0.12)	0.15 (0.05, 0.24)	-0.14 (-0.18, -0.09)
Social support	0.09 (0.03, 0.14)	0.002 (-0.05, 0.05)	0.01 (-0.04; 0.06)

Abbreviations: 95% CI, 95% Confidence Interval
Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". PU, process utility, is a derived measure representing the value a carer attaches to the process of informal caregiving.

Table 4 Multivariable multivariate analysis of CarerQol-VAS and Self Rated Burden (SRB)

	CarerQol-VAS				SRB			
	Unadjusted		Adjusted		Unadjusted		Adjusted	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Intercept	88.41	83.90, 92.92	92.17	86.67, 97.66	-1.86	-7.48, 3.76	-4.48	-12.04, 3.08
Fulfilment	2.34	1.38, 3.31	2.27	1.24, 3.29	-2.03	-3.25,-0.80	-2.76	-4.23, -1.29
Relational problems	-2.81	-3.80, -1.82	-2.38	-3.45, -1.32	8.97	7.71, 10.22	7.73	6.20, 9.26
Mental health issues	-7.96	-9.02, -6.89	-7.41	-8.57, -6.24	5.67	4.32, 7.03	5.48	3.81, 7.15
Physical health issues	-3.53	-4.57, -2.50	-1.37	-2.66, -0.09	3.82	2.49 5.14	2.97	1.13, 4.81
Problems with daily activities	-1.00	-2.08, 0.08	-1.47	-2.65, -0.29	11.04	9.68, 12.40	10.30	8.61, 12.00
Financial problems	-2.12	-3.93, -0.30	-1.00	-3.04, 1.04	2.15	-0.20, 4.49	2.49	-0.43, 5.42
Social support	0.58	-0.20, 1.36	0.19	-0.65, 1.03	0.89	-0.12, 1.90	0.25	-0.96, 1.46
Caregiver's sex (referent: female)			-1.07	-2.58, 0.44			-0.58	-2.74, 1.59
Caregiver's age			0.03	-0.03, 0.08			0.02	-0.06, 0.10
Socioeconomic status quartile								
Low			-0.41	-2.07, 1.24			1.23	-1.11, 3.57
Lower middle			0.01	-1.78, 1.79			-0.82	-3.35, 1.71
Upper middle			-1.48	-3.19, 0.24			-0.63	-3.07, 1.80
High			Reference				Reference	
Self-reported health								
Excellent/Very good			Reference				Reference	
Good			-4.59	-6.14, -3.04			-1.58	-3.80, 0.65
Fair/Poor			-6.89	-8.91, -4.87			-0.16	-3.07, 2.75
Care recipient's sex (Reference: female)			-1.05	-2.49, 0.38			1.47	-0.59, 3.53
Care recipient's age			0.11	0.02, 0.20			0.02	-0.11, 0.15
Care recipient's frailty index score			-9.38	-13.98, -4.78			31.50	24.91, 38.09

Cluster effects				
Cluster variance (σ_u^2)	1.83	4.37	70.46	84.59
Error variance (σ_e^2)	201.41	186.73	381.24	368.34
Intraclass correlation coefficient (ϱ)	0.009	0.02	0.16	0.19

Abbreviation: 95% CI, 95% Confidence Interval
Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". Unstandardised beta (β) estimates are presented. Models are adjusted for all variables listed on the table. The centred mean age was used in the multiple multivariate analysis.

Table 5 Interaction estimates for sampling frame in the multivariate analysis of CarerQol-VAS and Self-Rated Burden (SRB)

	CarerQol-VAS		SRB	
	β	95% CI	β	95% CI
Intercept	85.58	79.09, 92.07	6.39	-14.27, 0.51
Fulfilment	3.66	2.28, 5.06	-2.89	-4.51, -1.27
Relational problems	-2.49	-3.86, -1.12	9.21	7.61, 10.81
Mental health issues	-7.94	-9.41, -6.47	4.35	2.63, 6.07
Physical health issues	-2.90	-4.32, -1.48	3.93	2.25, 5.61
Problems with daily activities	-0.93	-2.43, 0.57	9.56	7.82, 11.31
Financial problems	-3.57	-6.22, -0.92	2.54	-0.61, 5.68
Social support	0.72	-0.38, 1.82	0.90	-0.42, 2.23
Sampling frame				
Primary care	Reference			
General	9.28	-0.38, 18.94	-15.93	-28.24, -3.62
Hospital	-0.22	-10.77, 10.32	-18.24	-32.06, 4.42
Fulfilment*				
General	-3.01	-5.25, -0.77	2.02	-0.92, 4.95
Hospital	-1.46	-3.99, 1.06	0.67	-2.71, 4.05
Relational problems*				
General	-1.35	-3.63, 0.94	0.56	-2.46, 3.58
Hospital	0.93	-1.78, 3.64	**-3.81**	**-7.48, -0.15**
Mental health issues*				
General	-0.38	-2.87, 2.11	2.54	-0.75, 5.83
Hospital	0.62	-2.25, 3.49	**4.15**	**0.21, 8.09**
Physical health issues*				
General	-1.33	-3.73, 1.08	-1.15	-4.34, 2.03
Hospital	-1.41	-4.25, 1.42	1.00	-2.86, 4.87
Problems with daily activities*				
General	-0.20	-2.70, 2.31	**3.56**	**0.26, 6.89**
Hospital	0.02	-2.86, 2.91	3.66	-0.23, 7.55
Financial problems*				
General	1.34	-3.09, 5.76	-0.96	-6.82, 4.90
Hospital	3.90	-0.48, 8.29	-0.01	-5.90, 5.87

Table 5 Continued

	CarerQol-VAS		SRB	
	β	95% CI	β	95% CI
Social support*				
General	0.72	-1.09, 2.54	-0.61	-3.02, 1.80
Hospital	-1.48	-3.52, 0.57	1.12	-1.65, 3.90
Cluster effects				
Cluster variance (σ_u^2)	1.70		75.72	
Error variance (σ_e^2)	200.83		379.79	
Intraclass correlation coefficient (ϱ)	0.008		0.17	

Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". Unstandardized beta (β) estimates are presented. Statistically significant interaction terms are indicated in bold. An asterisk (*) indicates an interaction term.

Table 6 Interaction estimates for survey mode in the multivariate analysis of CarerQol-VAS and Self-Rated Burden (SRB)

	CarerQol-VAS		SRB	
	β	95% CI	β	95% CI
Intercept	87.09	81.77, 92.42	-6.88	-14.27, 0.51
Fulfilment	2.51	1.35, 3.66	-2.34	-3.99, -0.69
Relational problems	-1.86	-3.07, -0.66	8.83	7.11, 10.54
Mental health issues	-8.51	-9.76, -7.26	5.77	3.98, 7.57
Physical health issues	-4.13	-5.38, -2.89	5.60	3.83, 7.38
Problems with daily activities	-0.20	-1.47, 1.06	11.49	9.69, 13.30
Financial problems	-2.45	-4.50, -0.39	4.15	1.21, 7.10
Social support	0.79	-0.14, 1.71	-0.01	-1.32, 1.30
Survey mode (questionnaire=reference)	-8.78	-21.38, 3.82	55.97	38.17, 73.77
Fulfilment*Mode	2.09	-0.99, 5.17	-1.81	-6.17, 2.54
Relational problems*Mode	-2.13	-4.87, 0.61	-2.79	-6.69, 1.10
Mental health issues*Mode	1.18	-2.03, 4.39	-3.07	-7.64, 1.49
Physical health issues*Mode	**4.82**	**1.80, 7.84**	**-9.35**	**-13.64, -5.07**

Table 6 Continued

	CarerQol-VAS		SRB	
	β	95% CI	β	95% CI
Problems with daily activities*Mode	-2.97	-6.23, 0.30	**-7.62**	**-12.26, -2.98**
Financial problems*Mode	0.15	-5.76, 6.05	-7.89	-16.31, 0.53
Social support *Mode	-0.30	-2.70, 2.11	3.34	-0.09, 6.76
Cluster effects				
Cluster variance (σ_u^2)	2.77		80.43	
Error variance (σ_e^2)	190.85		372.33	
Intraclass correlation coefficient (ϱ)	0.01		0.18	

Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". Statistically significant interaction terms are indicated in bold. Mode was defined as either interview or written questionnaire (reference group). An asterisk (*) indicates an interaction term.

Discussion

Our examination of the CarerQol instrument in different study settings demonstrates that construct validity was maintained. Although good scientific practice emphasizes the importance of re-evaluating instrument properties in individual research studies[15], our findings lend support to the applicability of the CarerQol instrument in a variety of settings. Nonetheless, pooled analyses based on the full TOPICS-MDS may not be appropriate for all research questions due to minor differential reporting. Survey mode in particular seemed to modestly influence self-reported burden. Given the available data, we are unable to discern why such differences arose, and we cannot exclude that some findings may be due to statistical chance alone. Still, different survey administration modes have been shown to influence response patterns[12,13,30], and caregivers may have underreported certain difficulties to frame their responses as more "socially desirable". Importantly, these findings should encourage future users of the CarerQol-7D and TOPICS-MDS to be mindful that certain outcome measures, such as SRB, may be more heavily influenced by study design features than others.

In our present study, we expanded on previous research by validating the CarerQol among caregivers who were sampled from the general population, hospitals and primary care centres. Reassuringly, observed associations between the CarerQol dimensions and outcomes of interest were similar.[16,17] However, complementary research examining other care settings, such as retirement communities or nursing homes, would also be of benefit. Such settings are unique given the integration of informal and formal care. TOPICS-MDS did contain two studies sampled from nursing homes; though, these studies were excluded due to small numbers (100 observations in total).

Although we reported the distributional properties of Transfer, it is important to emphasize that this VAS was developed in conjunction with the CarerQol-VAS with the intention of measuring PU. Transfer has not been independently validated for use in health services research, and thus should not be used as an individual marker of care-related burden. Nonetheless, we felt it was informative to present the distributional properties of Transfer since this measures directly influences the range of PU scores. If differential reporting was observed for one VAS and not the other, then the PU measurement may need to be re-evaluated. However, we found that neither VAS seemed to be disproportionately influenced by sampling frame or survey mode.

Our work is subject to several limitations. Firstly, the operationalization of concepts may raise concerns, specifically how caregivers interpret the CarerQol-VAS (happiness) and SRB (burden). However, using a VAS to measure general happiness has been widely applied in psychological and economic research[31] as well as specifically in the context of informal care.[32,33] Similarly, when compared with different burden scales, SRB produced a valid measure of burden.[21] Furthermore, the lack of a universally accepted operational definition for informal care may impede the applicability of findings. TOPICS-MDS applied a broad definition, defining informal care as long-term, voluntary, unpaid care for individuals with limitations impeding their ability to meet their daily needs. This definition did not specify a time duration for "long-term" care, though recent guidelines developed in the Netherlands have defined "long-term" as care provision lasting at least two weeks.[34]

Secondly, due to extensive missingness in the survey mode, we were inhibited from stratifying our analysis by both sampling frame and survey mode. Since the majority of interviews were conducted on caregivers sampled from primary care centers, we were unable to ascertain the interaction between these two study design features. Further stratification may have exposed additional complexities in the dataset. It is also important to note that survey mode was determined according to individual study protocols. The majority of included studies used written questionnaires, three were mixed mode and one

was interview only. To reduce bias, data collection was performed using standardized forms, and we accounted for clustering effects within individual studies by performing linear mixed models.

Lastly, our current analysis revealed minor differential reporting by survey mode; we have not discussed treatment of differential reporting in TOPICS-MDS. Since the magnitude and directionality of bias may vary across individual survey items, quantifying the sole impact of a study design feature is challenging. We speculate that uniform solutions are unlikely given that differential reporting is dependent upon the outcomes and covariates of interest. We do, however, encourage future users of TOPICS-MDS to apply meta-analytic techniques to examine heterogeneity and assess the feasibility of using pooled data.[35] Such techniques may include the *a priori* decision to perform subgroup or sensitivity analyses by survey mode.[36] Whereas random effects meta-analyses may be suitable for certain research questions, it is critical to emphasize that such analyses do not "fix" issues of heterogeneity.[36]

In this regard, a major strength of our study was that we were able to perform an individual patient data meta-analysis using TOPICS-MDS, a standardized, pooled master database. There are numerous advantages of using individual patient data over aggregate data, such as the ability to include unpublished data (thus reducing publication bias) and the ability to perform standardized statistical analyses across studies.[37] As data sharing becomes increasing encouraged in the scientific community[38], attempted retrieval of individual patient data should be encouraged in the protocol phase of a meta-analysis. Although individual patient data meta-analysis cannot circumvent all the challenges associated with pooled data[39,40], it can enhance researchers' understanding of the data and the effects of heterogeneity.

An additional strength is that we defined measurement properties (i.e. validity) in line with previous work published by COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Initiative.[26] Variation in terminology and definitions for measurement properties frequently occur in scientific research. By adhering to recommended terms and definitions reached through international consensus, we aim to create a greater transparency in our work.

In conclusion, our analysis supports the overall validity of the CarerQol instrument. This finding is not only relevant for individuals who wish to access TOPICS-MDS data, but also individuals who would like to apply the CarerQol instrument in future studies of care-related quality of life. Due to minor differential reporting, pooling mixed mode CarerQol data should be interpreted with caution; for TOPICS-MDS, meta-analytic techniques may be warranted.

References

1. Robine JM, Michel JP, Herrmann FR. Who will care for the oldest people in our ageing society? *BMJ*. Mar 17 2007;334(7593):570-571.
2. Agree EM, Glaser K. Demography of Informal Caregiving. In: Uhlenberg P, ed. *International Handbook of Population Aging*: Springer Netherlands; 2009:647-668.
3. Ward BW, Schiller JS. Prevalence of multiple chronic conditions among US adults: estimates from the National Health Interview Survey, 2010. *Prev Chronic Dis*. 2013;10:E65.
4. Caughey GE, Vitry AI, Gilbert AL, Roughead EE. Prevalence of comorbidity of chronic diseases in Australia. *BMC Public Health*. 2008;8:221.
5. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. *Eur J Gen Pract*. 2008;14 Suppl 1:28-32.
6. Bevans M, Sternberg EM. Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. *JAMA*. Jan 25 2012;307(4):398-403.
7. Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. *JAMA*. Dec 15 1999;282(23):2215-2219.
8. Joling KJ, van Hout HP, Schellevis FG, et al. Incidence of depression and anxiety in the spouses of patients with dementia: a naturalistic cohort study of recorded morbidity with a 6-year follow-up. *Am J Geriatr Psychiatry*. Feb 2010;18(2):146-153.
9. Bobinac A, van Exel NJ, Rutten FF, Brouwer WB. Health effects in significant others: separating family and care-giving effects. *Med Decis Making*. Mar-Apr 2011;31(2):292-298.
10. Mestheneos E, Triantafillou J (Eds.). *Supporting Family Carers of Older People in Europe: The Pan-European Background Report. Empirical Evidence, Policy Trends and Future Perspectives*. Hamburg: Lit Verlag2005.
11. Lutomski JE, Baars MA, Schalk BW, et al. The Development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): A Large-Scale Data Sharing Initiative. *PLoS One*. 2013;8(12):e81673.
12. Martin P, Lynn P. The effects of mixed mode survey designs on simple and complex analyses; Centre for Comparative Social Surveys Working Paper Series: Paper no. 04. Centre for Comparative Social Surveys, City University London; 2011.
13. Hood K, Robling M, Ingledew D, et al. Mode of data elicitation, acquisition and response to surveys: a systematic review. *Health Technol Assess*. May 2012;16(27):1-162.
14. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Q*. Sum 2008;72(2):167-189.
15. Streiner DL, Norman GR. *Health Measurement Scales: A pratical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008:167-210; 247-276.
16. Hoefman RJ, van Exel NJ, Foets M, Brouwer WB. Sustained informal care: the feasibility, construct validity and test-retest reliability of the CarerQol-instrument to measure the impact of informal care in long-term care. *Aging Ment Health*. Nov 2011;15(8):1018-1027.
17. Hoefman RJ, van Exel NJ, Looren de Jong S, Redekop WK, Brouwer WB. A new test of the construct validity of the CarerQol instrument: measuring the impact of informal care giving. *Qual Life Res*. Aug 2011;20(6):875-887.
18. Brouwer WB, van Exel NJ, van Gorp B, Redekop WK. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Qual Life Res*. Aug 2006;15(6):1005-1021.
19. Hoefman RJ, van Exel J, Rose JM, van de Wetering EJ, Brouwer WB. A discrete choice experiment to obtain a tariff for valuing informal care situations measured with the CarerQol instrument. *Med Decis Making*. Jan 2014;34(1):84-96.
20. Hoefman RJ, van Exel J, Brouwer W. How to include informal care in economic evaluations. *Pharmacoeconomics*. Dec 2013;31(12):1105-1119.

21. van Exel NJ, Scholte op Reimer WJ, Brouwer WB, van den Berg B, Koopmanschap MA, van den Bos GA. Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: a comparison of CSI, CRA, SCQ and self-rated burden. *Clin Rehabil*. Mar 2004;18(2):203-214.
22. Brouwer WB, van Exel NJ, van den Berg B, van den Bos GA, Koopmanschap MA. Process utility from providing informal care: the benefit of caring. *Health Policy*. Sep 28 2005;74(1):85-99.
23. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8:24.
24. Lutomski JE, Baars MA, van Kempen JA, et al. Validation of a frailty index from the older persons and informal caregivers survey minimum data set. *J Am Geriatr Soc*. Sep 2013;61(9):1625-1627.
25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. Jan 2007;60(1):34-42.
26. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. Jul 2010;63(7):737-745.
27. Schwarzer G. 2013; *meta*: Meta-analysis with R, http://cran.r-project.org/web/packages/meta/meta.pdf, Version = 3.1-2
28. Hopkins WG. A new view of statistics: Effect magnitudes. [Retrieved from http://www.sportsci.org/resource/stats/effectmag.html on 27 February 2014]. 2002.
29. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Stat*. Win 1998;23(4):323-355.
30. De Leeuw ED. To mix or not to mix data collection modes in surveys. *J Official Stat*. 2005;21:233-255.
31. van Praag BMS, Frijters P, Ferrer-i-Carbonell A. The anatomy of subjective well-being. *J Econ Behav Organ*. 2003;51:29-49.
32. Bobinac A, van Exel NJA, Rutten FFH, Brouwer WBF. Caring for and caring about: Disentangling the caregiver effect and the family effect. *J Health Econ*. 2010;29:549-556.
33. van den Berg B, Ferrer-i-Carbonell A. Monetary valuation of informal care: The well-being valuation method. *Health Econ*. 2007;16:1227-1244.
34. Hoefman RJ, Van Exel NJA, Brouwer WBF. iMTA Valuation of Informal Care Questionnaire (iVICQ). Version 1.0 (December 2011). Rotterdam: iBMG/iMTA, 2011. [Retrieved from www.bmg.eur.nl/english/imta/publications/manuals_questionnaires/ on 25 September 2013].
35. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.
36. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from www.cochrane-handbook.org; 2009.
37. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.
38. Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *Lancet*. May 14 2011;377(9778):1633-1635.
39. Riley RD. Commentary: like it and lump it? Meta-analysis using individual participant data. *Int J Epidemiol*. Oct 2010;39(5):1359-1361.
40. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*. 2012;344:d7762

Supplemental Table 1 Hypotheses on magnitude and directionality of correlations between CarerQol dimensions and CarerQol-VAS, Self-Rated Burden and Process Utility

	CarerQol-VAS	SRB	PU
SRB	Moderate, -	--	--
PU	Strong, +	Moderate, -	--
Fulfilment	Weak, +	Weak, -	Moderate, +
Relational problems	Moderate, -	Moderate, +	Moderate, -
Mental health issues	Moderate-Strong, -	Moderate, +	Moderate, -
Physical health issues	Moderate, -	Moderate, +	Moderate, -
Problems with daily activities	Moderate, -	Strong, +	Moderate, -
Financial problems	Weak, -	Weak, +	Weak, -
Social support	Trivial, +	Trivial, -	Weak, +

Abbreviations: SRB, Self-rated burden; PU, Process Utility **Note:** Correlation coefficients were classified as trivial (<0.1), weak (0.1 to <0.3), moderate (0.3 to <0.5), strong (0.5 to <0.7), very strong (>0.7). A '+' signs indicates a hypothesized positive correlation whereas a '-' indicates a hypothesized negative correlation.

Supplemental Table 2 Random effects meta-correlation coefficient (Spearman's rho) by sampling frame

	CarerQol-VAS (95% CI)	SRB (95% CI)	PU (95% CI)
General population			
Burden	-0.36 (-0.42, -0.29)	--	--
Process	0.49 (0.43, 0.54)	-0.33 (-0.43, -0.23)	--
CarerQol domains			
Fulfilment	0.15 (0.03, 0.26)	-0.15 (-0.30, 0.01)	0.22 (0.12, 0.31)
Relational problems	-0.37 (-0.52, -0.21)	0.44 (0.34, 0.53)	-0.27 (-0.35, -0.19)
Mental health issues	-0.45 (-0.54, -0.34)	0.39 (0.32, 0.45)	-0.27 (-0.34, -0.20)
Physical health issues	-0.33 (-0.40, -0.25)	0.34 (0.21, 0.45)	-0.26 (-0.41, -0.11)
Problems with daily activities	-0.24 (-0.35, -0.12)	0.47 (0.37, 0.56)	-0.25 (-0.38, -0.10)
Financial problems	-0.14 (-0.21, -0.06)	0.17 (0.10, 0.24)	-0.10 (-0.17, -0.03)
Social support	0.03 (-0.08, 0.14)	0.05 (-0.06, 0.15)	<0.001 (-0.12, 0.12)

Supplemental Table 2 Continued

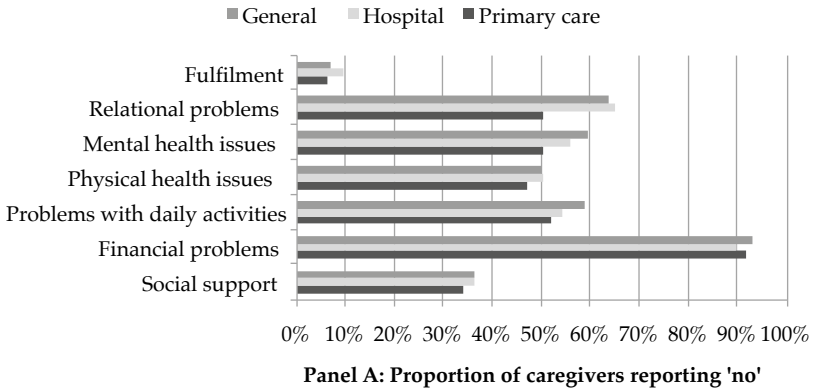
	CarerQol-VAS (95% CI)	SRB (95% CI)	PU (95% CI)
Hospital			
Burden	-0.30 (-0.38, -0.21)	--	--
Process	0.56 (0.50, 0.62)	-0.43 (-0.50, -0.35)	--
CarerQol domains			
Fulfilment	0.16 (0.07, 0.24)	-0.18 (-0.29, 0.06)	0.20 (0.12, 0.29)
Relational problems	-0.22 (-0.35, -0.07)	0.36 (0.28, 0.43)	-0.29 (-0.37, -0.20)
Mental health issues	-0.36 (-0.48, -0.22)	0.43 (0.36, 0.50)	-0.29 (-0.37, 0.21)
Physical health issues	-0.29 (-0.41, -0.16)	0.40 (0.32, 0.47)	-0.32 (-0.44, -0.18)
Problems with daily activities	-0.27 (-0.35, -0.18)	0.47 (0.36, 0.56)	-0.36 (-0.44, -0.28)
Financial problems	-0.10 (-0.20, 0.01)	0.13 (-0.05, 0.30)	-0.09 (-0.23, 0.06)
Social support	0.03 (-0.09, 0.15)	0.05 (-0.04, 0.14)	-0.04 (-0.16, 0.09)
Primary care			
Burden	-0.17 (-0.47, 0.17)	--	--
Process	0.51 (0.47, 0.55)	-0.27 (-0.46, -0.04)	--
CarerQol domains			
Fulfilment	0.23 (0.18, 0.28)	-0.15 (-0.25, -0.04)	0.35 (0.27, 0.42)
Relational problems	-0.27 (-0.31, -0.21)	0.30 (0.14, 0.45)	-0.33 (-0.38, -0.28)
Mental health issues	-0.46 (-0.51, -0.40)	0.28 (0.09, 0.44)	-0.30 (-0.36, -0.23)
Physical health issues	-0.37 (-0.42, -0.32)	0.29 (0.07, 0.48)	-0.24 (-0.29, -0.19)
Problems with daily activities	-0.27 (-0.32, -0.22)	0.36 (0.13, 0.55)	-0.33 (-0.39, -0.28)
Financial problems	-0.21 (-0.26, -0.16)	0.13 (-0.01, 0.27)	-0.16 (-0.22, -0.11)
Social support	0.14 (0.09, 0.20)	-0.02 (-0.08, 0.05)	0.05 (-0.002, 0.11)

Abbreviation: 95% CI, 95% Confidence Interval
Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". PU, process utility, is a derived measure representing the value a carer attaches to the process of informal caregiving.

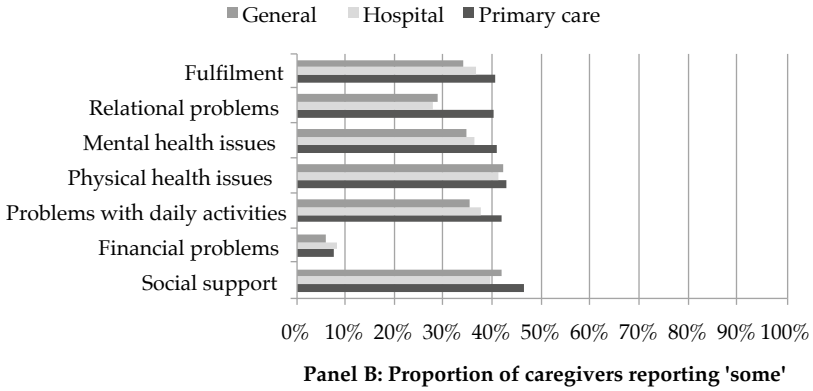
Supplemental Table 3 Random effects meta-correlation coefficient (Spearman's rho) by survey mode

	CarerQol-VAS (95% CI)	SRB (95% CI)	PU (95% CI)
Questionnaire			
Burden	-0.34 (-0.39, -0.29)	--	--
Process	0.53 (0.50, 0.56)	-0.42 (-0.47, -0.37)	--
CarerQol domains			
Fulfilment	0.20 (0.16, 0.25)	-0.20 (-0.26, -0.15)	0.29 (0.22, 0.36)
Relational problems	-0.26 (-0.32, -0.21)	0.41 (0.37, 0.46)	-0.33 (-0.37, -0.28)
Mental health issues	-0.44 (-0.50, -0.38)	0.39 (0.35, 0.43)	-0.30 (-0.35, -0.25)
Physical health issues	-0.35 (-0.40, -0.30)	0.40 (0.35, 0.43)	-0.28 (-0.34, -0.21)
Problems with daily activities	-0.26 (-0.30, -0.21)	0.48 (0.43, 0.52)	-0.34 (-0.38, -0.29)
Financial problems	-0.18 (-0.22, -0.13)	0.20 (0.13, 0.26)	-0.15 (-0.20, -0.09)
Social support	0.10 (0.03, 0.16)	-0.01 (-0.08, 0.05)	0.01 (-0.06, 0.09)
Interview			
Burden	-0.12 (-0.72, 0.58)	--	--
Process	0.45 (0.36, 0.54)	-0.10 (-0.46, 0.30)	--
CarerQol domains			
Fulfilment	0.29 (0.17, 0.39)	0.01 (-0.27, 0.29)	0.22 (0.005, 0.41)
Relational problems	-0.28 (-0.44, -0.11)	0.13 (-0.26, 0.49)	-0.24 (-0.44, -0.03)
Mental health issues	-0.40 (-0.49, -0.30)	0.23 (-0.33, 0.67)	-0.28 (-0.38, -0.17)
Physical health issues	-0.26 (-0.49, -0.001)	0.10 (-0.46, 0.60)	-0.30 (-0.40, -0.19)
Problems with daily activities	-0.26 (-0.42, -0.09)	0.21 (-0.45, 0.72)	-0.26 (-0.43, -0.07)
Financial problems	-0.13 (-0.35, 0.10)	0.02 (-0.34, 0.37)	-0.11 (-0.28, 0.07)
Social support	0.10 (-0.01, 0.21)	0.10 (-0.02, 0.21)	0.01 (-0.15, 0.18)

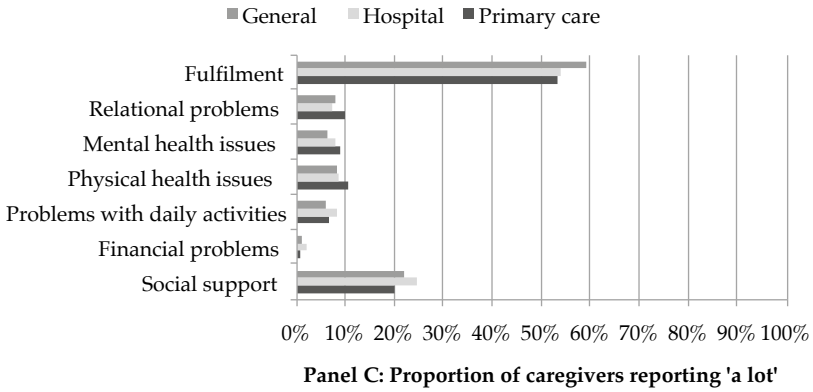
Abbreviation: 95% CI, 95% Confidence Interval
Note: CarerQol-VAS refers to the statement, "Please indicate how happy you are currently feeling". SRB refers to the statement, "Please indicate how burdensome you feeling care for loved one is at the moment". PU, process utility, is a derived measure representing the value a carer attaches to the process of informal caregiving.



Panel A: Proportion of caregivers reporting 'no'

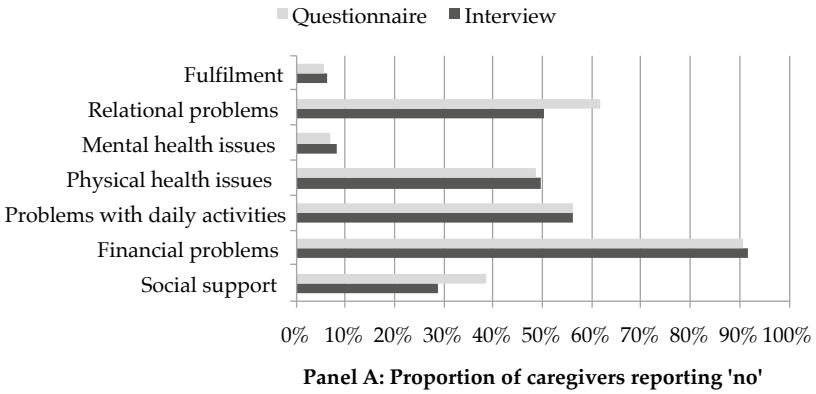


Panel B: Proportion of caregivers reporting 'some'

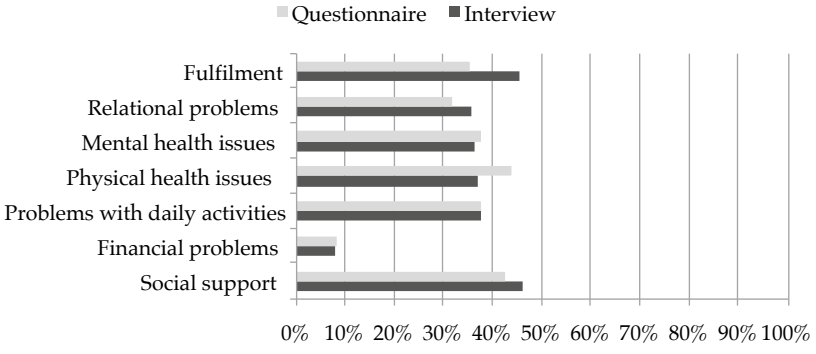


Panel C: Proportion of caregivers reporting 'a lot'

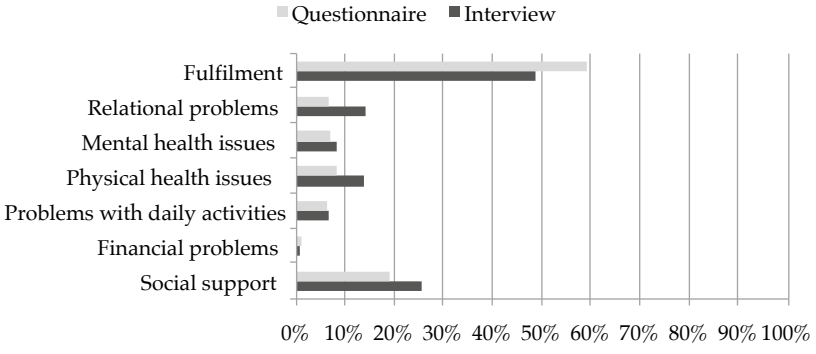
Supplemental Figure 1 Distribution of CarerQol-7D dimensions by sampling framework (General population N=753; Hospital N=576; Primary care N=1,940)



Panel A: Proportion of caregivers reporting 'no'



Panel B: Proportion of caregivers reporting 'some'



Panel C: Proportion of caregivers reporting 'a lot'

Supplemental Figure 2 Distribution of CarerQol-7D dimensions by survey mode (Questionnaire N=1,895; Interview N=300)

4

Validation of Variables Derived from TOPICS-MDS

4.1

Measuring frailty: Validating TOPICS Frailty Index

Published as:

Lutomski JE, Baars MA, van Kempen JA, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Krabbe PF, Steunenberg B, Steyerberg EW, Olde-Rikkert MG, Melis RJ. Validation of a frailty index from the older persons and informal caregivers survey minimum data set. *Journal of the American Geriatrics Society*. 2013 Sep;61(9):1625-7.

To the Editor:

Frailty refers to a state of increased vulnerability to adverse health outcomes.[1] Although clinically assessing a breadth of health complexities is a robust method to measure frailty[1], such a detailed assessment is timely and cost-prohibitive in large studies. Thus, a growing body of literature has begun to apply a frailty index (FI) based on the concept of 'deficit accumulation'.[2-5] Deficits in health may include a range of symptoms, morbidities or functional limitations.[6] Since such deficits can be easily extracted from survey data, we sought to assess construct validity between a FI derived from a minimal dataset versus a clinical assessment, our reference standard.

Methods

Data were derived from the Easycare Two-Step Older Persons Screening Study.[7] Patients ≥70 years were randomly sampled from six primary care practices in Nijmegen, Netherlands to test a new frailty identification tool. All participants received a clinical assessment and were also asked to complete The Older Persons and Informal Caregivers Survey Minimal DataSet (TOPICS-MDS) as part of a national initiative to create a minimal dataset on older persons' health. Details on this initiative are available elsewhere.[8] TOPICS-MDS collected information on morbidity status, functional limitations using a modified version of the Katz Activities of Daily Living (ADL) Index[9] and quality of life using the EuroQoL 5D+C instrument.[10]

We derived FIs from the clinical assessment and TOPICS-MDS using a validated method previously defined by Mitnitski et al.[2] In brief, the sum of deficits observed in a participant is divided by the total number of deficits under review to calculate a proportion. The Clinical-FI was constructed using 45 deficits. Since analyzing the statistical interaction between frailty, basic ADL, and morbidity status may be of future interest for researchers, we constructed a Long and Short TOPICS-FI using 46 and 23 deficits respectively. The Long TOPICS-FI comprised of morbidity status, basic and instrumental ADL, and quality of life, whereas the Short TOPICS-FI comprised of only instrumental ADL and quality of life.

To assess construct validity of the Long and Short TOPICS-FIs, we performed a Pearson's correlation with the Clinical-FI. We examined their distribution across sex and age group (70-74; 75-79; 80+ years) as these are established risk factors for frailty.[5] Significant differences in mean FI scores by sex and age group were determined using a t-test and ANOVA (SAS V9.2; Carey, NC).

Results

Of the 587 participants sampled, the majority had no missing data points for the FIs (Clinical: 92.2%; Long and Short TOPICS: 62.2%). The maximum FI scores recorded for the clinical assessment, Long and Short TOPICS were 0.58, 0.52 and 0.68 respectively. There was a moderate correlation between the Clinical-FI and the Long and Short TOPICS-FIs ($r=0.75$, $p<0.0001$; $r=0.73$, $p<0.0001$ respectively). Expectedly, the two TOPICS-FIs were highly correlated ($r=0.93$, $p<0.0001$). The overall distribution for both TOPICS-FIs were tailed to the right (data not shown), though became more normalized when stratified by age (Figure 1, results shown for Long TOPICS-FI only). Mean scores (±SD) significantly differed across sex and age for both the Long TOPICS-FI [(**Men:** 0.14(0.07); **Women:** 0.16(0.08); $p=0.003$) (**70-74:** 0.12(0.05); **75-79:** 0.15(0.06); **80+:** 0.21(0.09); $p<0.001$)] and the Short TOPICS-FI [(**Men:** 0.22(0.11); **Women:** 0.25(0.11); $p=0.013$) (**70-74:** 0.19(0.08); **75-79:** 0.23(0.10); **80+:** 0.31(0.12); $p<0.001$)].

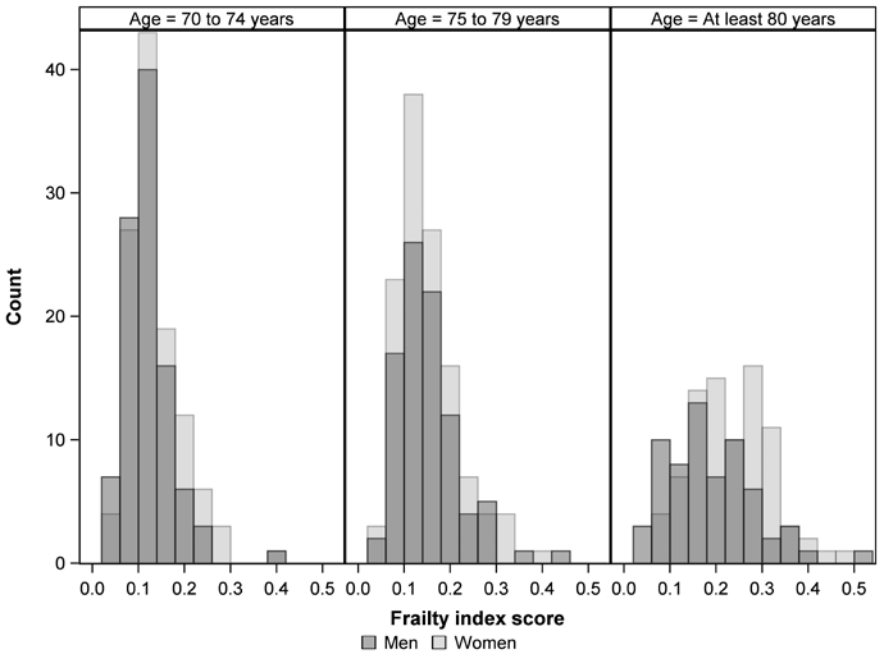


Figure 1 Frequency distribution of a 46-item frailty index derived from The Older Persons and Informal Caregivers Survey according to 5-year age groups and sex (N = 587)

Discussion

Our findings support that a FI derived from TOPICS-MDS can assess frailty with reasonable accuracy. Similar to previous research, the overall distribution of frailty was skewed and there were clear patterns by age and sex.[2-4] Our analysis further suggests that it seems possible to create a valid FI independent of morbidity and basic ADL, which would facilitate future analyses of their interaction. Moreover, since data points may be limited, yielding valid results using different base criteria is informative for researchers seeking to calculate a FI from secondary data sources.

There are several inherent advantages to using a short, standardized survey, like TOPICS-MDS, to derive a FI: ease of administration, cost-effectiveness, and efficiency (i.e. clinical review of case notes is not required). Still, given its sampling frame, construct validity of TOPICS-FIs should be examined in a broader population to ensure generalizability. Although out of the scope of this study, assessing the predictive validity of the Short and Long TOPICS-FIs for subsequent risk of mortality and/or other adverse outcomes would be of benefit. This would be particularly relevant for the Short TOPIC-FI, since indices based on fewer items may produce less stable estimates of frailty.[6]

Frailty has become a central topic in elderly health research. The ability to accurately identify frailty among older persons from minimal survey data can guide interventions leading to improved health outcomes.

References

1. Clegg A, Young J, Iliffe S, Olde Rikkert M, Rockwood K. Frailty in elderly people. *Lancet.* 2013;381(9868):752-762.
2. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Scientific World Journal.* Aug 8 2001;1:323-336.
3. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ.* Aug 30 2005;173(5):489-495.
4. Romero-Ortuno R, Kenny RA. The frailty index in Europeans: association with age and mortality. *Age Ageing.* Sep 2012;41(5):684-689.
5. Mitnitski AB, Mogilner AJ, MacKnight C, Rockwood K. The mortality rate as a function of accumulated deficits in a frailty index. *Mech Ageing Dev.* 2002;123(11):1457-1460.
6. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr.* 2008;8:24.
7. van Kempen JAL, Schers HJ, Jacobs A, et al. Development of an instrument for the identification of frail older people as a target population for integrated care. *Br J Gen Pract.* 2013;63(608):e225-e231.
8. The National Care for the Elderly Programme. Available at: http://www.nationaalprogramma-ouderenzorg.nl/english/the-national-care-for-the-elderly-programme/. Accessibility verified March 11, 2013.
9. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc.* Jun 1992;40(6):585-588.
10. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol.* Apr 1999;52(4): 293-301.

4.2

The interplay between frailty, activities of daily living, and multi-morbidity

Published as:

Lutomski JE, Baars MA, Boter H, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis R on behalf of TOPICS-MDS Consortium. Frailty, disability and multi-morbidity: the relationship with quality of life and healthcare costs in elderly people. *Nederlands Tijdschrift Geneeskunde*. 2014;158:A7297.

Abstract

Purpose: To assess the independent and combined impact of frailty, multimorbidity, and ADL limitations on self-reported quality of life and health care costs.

Setting: Cross-sectional interview data from The Older Persons and Informal Caregivers Minimum DataSet, a pooled dataset with information from 41 different research projects funded by the National Care for the Elderly Programme.

Methods: Analyses were stratified by living situation (independent or residential care facility). Directionality and magnitude of associations were assessed using linear mixed models, allowing for clustering effects between studies. Quality of life and health care costs were regressed on frailty, multimorbidity, ADL limitations, and the interactions between these domains.

Results: 11.093 older persons were interviewed. Frailty, multimorbidity, and ADL limitations were reported for a substantial proportion of older persons living independently (56.4%; 88.3%; 41.4% respectively) and older persons living in a residential care facility (88.7%; 89.2%; 77.3% respectively). One-third of older persons living at home (31.9%) reported all three conditions compared to two-thirds of older persons living in a residential care facility (68.3%). In the multivariate analysis, frailty had an independent, strong impact on outcomes. Irrespective of living situation, older persons experiencing problems across the three domains reported the poorest quality of life scores and the highest health care costs.

Conclusion: Frailty, multimorbidity, and ADL limitations are complementary measures which provide a more holistic understanding of health status. As such, taking this multi-dimensional approach in explaining quality of life and health care costs can help elucidate these complex relationships.

Introduction

Multimorbidity and limitations in Activities of Daily Living (ADL) are widely measured in clinical and public health research. However, as an increasing proportion of older persons present with multiple chronic conditions[1], such measurements may only provide a limited portrayal of the health and wellbeing. Thus, it is natural progression to move towards a broader, multidimensional model which encompasses the complexity of health states that health practitioners currently face.[2-4] In this context, measuring and monitoring frailty in older persons has been widely recognized as a useful indicator of the health status of older persons.[5]

However, to date, there is no general consensus on the precise definition or the optimal method of identification of frailty in older persons.[6] More narrow definitions focus on pathophysiological changes whereby older persons exposed to stressor have a predisposition to adverse health outcomes.[6] In contrast, broader definitions incorporate mental health and social functioning.[6-8] In this article, frailty was identified using a frailty index, which defines frailty as an accumulation of health deficits across multiple physiological domains.[7]

Undoubtedly, there is an underlying relationship between frailty, morbidity status, and disability, i.e. frail older persons are more likely to also have chronic co-morbidity and ADL limitations.[9,10] However, there has been limited research assessing the additional value of measuring frailty over solely multimorbidity or ADL limitations.[9,10] Thus, to demonstrate the interaction between frailty, multimorbidity, and ADL limitations, we assessed the independent and combined impact of these three domains on self-reported quality of life and health care costs.

Methods

Data source

Data were derived from TOPICS-MDS (www.topics-mds.eu), which is a public data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere.[11] Briefly, the Dutch National Care for the Elderly Programme (http://www.nationaal-programmaouderenzorg.nl) was established in 2008 to promote proactive, integrated health care for older persons with complex care needs. As part of this national agenda, TOPICS-MDS was developed to collect uniform information from all research projects funded under the Programme. Thus, TOPICS-MDS

consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardized protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to TOPICS-MDS, the database is dynamic in nature and thus regularly updated with new observations. The present analysis uses the first version of the dataset available as of January 2013 and is based on 41 research projects. TOPICS-MDS is a fully anonymised dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2012/120).

Definitions

Multimorbidity was measured using an adapted version of the *Lokale en Nationale Monitor Gezondheid* (i.e. a health monitoring survey) to identify 17 chronic diseases. Older persons were defined as multimorbid if they reported having two or more chronic conditions.[12]

ADL limitations were identified using the Katz Index of Independence in Activities of Daily Living.[13] This index assesses six basic functions: bathing, dressing, toileting, continence products, transfer, and eating. Older persons were identified as having an ADL limitation if they reported having at least one of these six limitations.

To assess the additional value of measuring frailty, we constructed a frailty index.[14] The frailty index is one of several methods widely used in population studies to estimate frailty and has been validated for use in TOPICS-MDS.[15] Briefly, a frailty index is calculated by reviewing health problems affecting different physiological systems; the total number of problems observed in an individual is then divided by the total number of problems reviewed to calculate a proportion (e.g. 10 problems observed /40 problem reviewed = 0.25).[16] For the purposes of this analysis, we defined frailty as a frailty index score $\geq$0.25. Using this standardized method, we constructed a 23-item frailty index using TOPICS-MDS data available on instrumental ADLs, quality of life, psychological health, self-perceived health status, and social functioning. Further details on how these indicators were measured have been described elsewhere.[11]

The first outcome of interest, self-perceived quality of life, was assessed with a modified version of Cantril's Self Anchoring Ladder.[17] Older persons were asked to rate their current quality of life on a scale between zero (worst possible state) and ten (best possible state).

The second outcome of interest, overall health care costs, comprised of hospitalizations, emergency GP care, home care, day care, and temporary/permanent stay in a care or nursing home reported in the 12 months prior to assessment.[18] Using a top-down costing method, standard costing rates were applied to each service the older person reported.

Analyses

Given that measurement bias is known to vary by survey mode[19], it was decided *a priori* to analyze interview data and written questionnaire data separately. Results shown in this paper are based on interview data; results based on written questionnaire data are available as web supplement. Analyses were further stratified by living situation (i.e. community-dwelling or residential care facility) due to different social dynamics and baseline differences in cost in these settings. Persons residing in nursing homes were excluded due to small numbers. Descriptive data were derived. Based on a complete case analysis, directionality and magnitude of associations were assessed using linear mixed models, which allowed for clustering effects between studies. Quality of life (Model 1) and health care costs (Model 2) were regressed on frailty, multimorbidity, ADL limitations, and the interactions between these domains. In both models, adjustments were made for the age, sex, and educational level. Statistical analyses were carried out using SPSS (Version 21.0. Armonk, NY: IBM Corp).

Results

11,093 older persons were interviewed for TOPICS-MDS with an average age of 80 years (standard deviation 6 years) (Table 1). Approximately one-third (36.3%) were men, one-fifth (18.3%) had pre-university training or higher and nearly half (46.4%) were widowed. Compared to persons living independently, older persons living in a residential care facility were older and more likely to be female, have lower education, be widowed, and have poorer self-reported health. Frailty, multimorbidity, and ADL limitations were reported for a substantial proportion of older persons living independently (56.4%; 88.3%; 41.4% respectively). Proportions were notably higher among older persons living in a residential care facility (88.7%; 89.2%; 77.3% respectively).

Older persons living independently were more likely to exclusively report having multimorbidity (i.e. no observed frailty or ADL limitations) compared to older persons living in a residential care facility (26.6% versus 5.0% respectively, Table 2). Whereas one-third of older persons living at home (31.9%) reported frailty, multimorbidity, and ADL limitations, two-thirds of older persons living

Table 1 Sociodemographic characteristics of participants who were interviewed by living situation, TOPICS-MDS, 2014

	Total (n=11,093)	Community-dwelling (n=9,622)	Residential care facility (n=1,471)
		Percentage	
Men	36.3	38.0	24.8
Education			
Primary school or lower	37.5	35.4	50.5
Secondary	44.2	44.9	39.3
Some college / College	18.3	19.6	10.2
Marital status			
Married / Cohabiting	42.1	45.9	17.2
Divorced / Single	11.5	11.2	13.2
Widowed	46.4	42.9	69.6
Nationality			
Native Dutch	90.7	90.1	94.5
First generation	5.5	5.8	3.7
Second generation	3.8	4.1	1.8
Self-reported health			
Excellent / Very good	9.7	10.3	5.2
Good	48.6	50.8	34.6
Fair / Poor	41.7	38.9	60.2
Self-reported health relative to previous year			
Much better/Somewhat better	9.5	9.4	9.8
Same	46.9	48.2	38.6
Somewhat worse/Much worse	43.6	42.4	51.6
Frailty	60.3	56.4	88.7
Multimorbidity	88.4	88.3	89.2
ADL limitations	46.2	41.4	77.3
		Average (SD)	
Age (years)	80 (6)	79 (6)	86 (6)
Quality of life	0.6 (0.3)	0.6 (0.3)	0.6 (0.3)
Psychological wellbeing	74.0 (18.2)	74.4 (17.9)	71.1 (19.4)

Note: Frailty was defined as a score ≥0.25 on 23-item frailty index derived from TOPICS-MDS. Multimorbidity was defined as the presence of ≥2 chronic conditions. Limitations in ADL were defined as requiring assistance with ≥1 item on the 6-item Katz Activities of Daily Living Index. Quality of life as measured using the EQ-5D summary score (score range -0.33 to 1.0, higher is better). Psychological wellbeing was measured using the Rand-36 Mini-Mental Health Scale (score range 0-100, higher is better).

in a residential care facility (68.3%) reported all three conditions. Thus, older persons living in a residential care facility were less likely to have one or two conditions and more likely to have all three conditions.

In the mixed model examining quality of life, older persons living independently who did not report multimorbidity, ADL limitations or frailty were the reference group; their average quality of life score was 7.5 (95% CI: 6.7; 8.3) (Table 3). Frailty had the largest independent impact on quality of scores [-0.9 points (95% CI: -1.1; -0.7) less than reference score, i.e. 6.6 average score] whereas limitations in ADLs had the least independent impact [0.03 points (95% CI: -0.3; 0.3)]. However, older persons reporting frailty, multimorbidity, and ADL limitations had the lowest quality of life scores; on average their score was -1.2 points (95% CI: -1.3; -1.1) less than the reference score. Similar patterns were observed for older persons living in a residential care facility.

Table 2 The distribution of frailty, multimorbidity and ADL limitations in participants who completed an interview by living situation

	Total (n=11,093) %	**Community-dwelling (n=9,622)** %	**Residential care facility (n=1,471)** %
No complications	6.9	7.4	3.0
Frailty only	1.7	1.7	1.9
Multimorbidity only	23.9	26.6	5.0
ADL limitations only	0.6	0.6	0.6
Multimorbidity and ADL limitations	7.0	7.6	2.8
Frailty and multimorbidity	21.4	22.5	13.6
Frailty and ADL limitations	2.0	1.6	4.8
Frailty, multimorbidity and ADL limitations	36.4	31.9	68.3
Total	**100%**	**100%**	**100%**

Note: Categories are mutually exclusive. Frailty was defined as a score ≥0.25 on 23-item frailty index derived from TOPICS-MDS. Multimorbidity was defined as the presence of ≥2 chronic conditions. Limitations in ADL were defined as a participant requiring assistance with at least one item on the 6-item Katz Activities of Daily Living Index.

Table 3 The association between self-reported quality of life scores and frailty, multimorbidity, and ADL limitations by living situation (based on interview data only)

	Community-dwelling	Residential care facility
No complications (reference group)	7.5 (6.7; 8.3)	7.7 (6.7; 8.7)
Frailty only	-0.9 (-1.1; -0.7)	-0.5 (-1.3; 0.3)
Multimorbidity only	-0.3 (-0.4; -0.2)	-0.3 (-0.8; 0.3)
ADL limitations only	0.03 (-0.3; 0.3)	-0.6 (-1.6; 0.5)
Multimorbidity and ADL limitations	-0.3 (-0.5; -0.2)	-0.2 (-0.9; 0.4)
Frailty and multimorbidity	-1.1 (-1.2; -0.9)	-1.1 (-1.6; -0.6)
Frailty and ADL limitations	-0.8 (-1.0; -0.6)	-0.6 (-1.2; -0.04)
Frailty, multimorbidity and ADL limitations	-1.2 (-1.3; -1.1)	-1.2 (-1.6; -0.7)

Note: Quality of life is rated from zero to ten, with higher scores represented higher quality of life. Results show the average difference (95% confidence interval) relative to the reference group. The model has been corrected for age, sex, education, and random effects by project.

In the mixed model examining health care costs, older persons living independently who did not report multimorbidity, ADL limitations or frailty were the reference group; their average health care costs per annum were €2,018 (95% CI: -€427; €4,464) (Table 4). ADL limitations had the largest independent impact on health care costs. On average, older persons living independently who only reported limitations in ADLs spent €578 (95% CI: -€838; €1,993)] more on health-related services. Multimorbidity had the least independent impact [€288 (95% CI: -€181; €758)]. Older persons with all three conditions (frailty, multimorbidity, and ADL limitations) reported the highest health care costs; on average they spent €2,772 (95% CI: €2,301; €3,243) more than the reference group.

For older persons living in a residential care facility, the average cost of care per annum was largely driven by residence costs (approximately €30,000 per annum). In contrast to older persons living independently, frailty had the largest independent impact on health care costs among older persons living in a residential care facility. On average, frail older persons spend €2,220 (95% CI: -€1,831; €6,271) more than individuals without any of these conditions. Similarly to older persons living independently, older persons living in a residential care facility with all three conditions reported the highest health care costs.

Table 4 The association between care costs (€) and frailty, multimorbidity, and ADL limitations by living situation (based on interview data only)

	Community-dwelling	Residential care facility
No complications (reference group)	2,018 (-427; 4,464)	32,631 (29,696; 35,566)
Frailty only	346 (-587; 1,280)	2,220 (-1,831; 6,271)
Multimorbidity only	288 (-181; 758)	-270 (-3,472; 2,932)
ADL limitations only	578 (-838; 1,993)	286 (-5,740; 6,311)
Multimorbidity and ADL limitations	694 (111; 1,276)	292 (-2,311; 4,978)
Frailty and multimorbidity	1,326 (844; 1,808)	627 (-2,178; 3,430)
Frailty and ADL limitations	1,081 (109; 2,053)	1,244 (-1,992; 4,480)
Frailty, multimorbidity and ADL limitations	2,772 (2,301; 3,243)	2,629 (21; 5,237)

Note: Results show the average difference (95% confidence interval) in care costs relative to the reference group. The model has been corrected for age, sex, education and random effects by project.

Patterns in quality of life scores were broadly similar among older persons who completed a written questionnaire. However, some costing differences were observed in this subgroup. Data on older persons who completed a questionnaire are available as web supplements (Tables S1-S4).

Discussion

This study supports that measuring the combined influence of frailty, multimorbidity, and ADL limitations is advantageous in understanding quality of life and health services utilization. By examining the independent effects and the interactions between these three conditions, it is evident that frailty serves as in important independent contributor to health outcomes, and this finding is not unique to this study. Frailty has been shown to represent a distinct condition separate from the culmination of morbidities or ADL limitations[9] and to be independently associated with increased risk of functional dependency[10], falls[20], institutionalization[21], and death.[14] Such adverse outcomes may be due to diminished autonomy and self-management in frail individuals leading to less successful aging.[22] Thus, although multimorbidity and ADL limitations are more frequently monitored, such findings should encourage clinicians, public

health practitioners and health policy makers to develop a framework to more regularly monitor frailty.

Nonetheless, this study is subject to several inherent limitations. Firstly, there are two prevailing approaches to identify frailty: a narrow approach which mainly focuses on physical vulnerability (e.g. the phenotype model) and broad-based approach (e.g. the Tilburg Frailty Indicator and the frailty index used in this study). The former defines frailty using five physiological indicators[3] whereas the latter typically includes both physiological and general well-being indicators.[25] Given the structure of TOPICS-MDS, only the cumulative deficit model (i.e. frailty index) could be applied. Yet, despite distinct approaches in measuring frailty, it is important to note that there is a considerable degree of overlap and statistical convergence between them.[6,23] Moreover, compared to the phenotype model, the cumulative deficit model may better differentiate between moderate and severe levels of frailty.[26]

Secondly, frailty indices are often based on at least 30 health criteria[16], and thus the frailty index applied in this study is relatively short and covers a limited number of domains. Reassuringly, however, the frailty index used in this analysis displayed similar distributional properties and was correlated with a longer (46-item) frailty index as well as frailty identified through clinical assessment.[15] Although such findings support the validity of the frailty index used in this study, further research is needed to assess the predicative validity of a 23-item frailty index to ensure its robustness in longitudinal research. Currently, only cross-sectional data are available in TOPICS-MDS. However, given the dynamic nature of the database, follow-up data will be available in late 2014 to investigate this issue.

Lastly, TOPICS-MDS is a pooled dataset of research projects with different sampling protocols and certain oversampled sub-groups (e.g. frail older persons). Thus, the prevalence of frailty, multimorbidity, and ADL limitations reported in this analysis are not reflective of the overall Dutch population. For this reason, it was unsurprising that frailty in our study population was higher than estimates based on the general population.[7] Investigations are underway to determine if demographic weighting can be applied to the database to facilitate health policy and resource planning. Still, the associations identified in this study should not be underestimated.[27] The multivariable model not only accounted for differences in the prevalence of frailty, multimorbidity, and ADL limitations but also key demographic characteristics (sex, age, and educational status and clustering effects between studies). Moreover, when the data were restricted to older persons sampled from the general population, similar associations were observed (data not shown). In light of research in other settings which has highlighted the combined impact of frailty, multimorbidity,

and ADL limitations[9], it is highly probable that associations observed in this paper are relevant for the Dutch population as well as other high-income countries.

In conclusion, these findings support a clear advantage for using a multi-dimensional approach in understanding quality of life and health care costs of older people. Frailty, multimorbidity, and ADL limitations should each be viewed as complementary measures which provide a more holistic understanding of health status. To further elucidate the complex relationship between frailty, multimorbidity, and ADL limitations on quality of life and health services costs, additional longitudinal studies would be of benefit.

References

1. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. *Lancet*. Oct 3 2009;374(9696):1196-1208.
2. Rockwood K, Fox RA, Stolee P, Robertson D, Beattie BL. Frailty in elderly people: an evolving concept. *CMAJ*. Feb 15 1994;150(4):489-495.
3. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. Mar 2001;56(3):M146-156.
4. Iliffe S. Family medicine in an aging society. *Can Fam Physician*. May 2009;55(5):463-467.
5. Sternberg SA, Wershof Schwartz A, Karunananthan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *J Am Geriatr Soc*. Nov 2011;59(11):2129-2138.
6. Theou O, Brothers TD, Pena FG, Mitnitski A, Rockwood K. Identifying Common Characteristics of Frailty Across Seven Scales. *J Am Geriatr Soc*. Apr 2 2014.
7. van Campen C. Kwetsbare ouderen. Den Haag: Sociaal en Cultureel Planbureau, Den Haag. Web. Accessed 9 April 2014 <http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2011/Kwetsbare_ouderen> 2011.
8. Cesari M, Gambassi G, van Kan GA, Vellas B. The frailty phenotype and the frailty index: different instruments for different purposes. *Age Ageing*. Jan 2014;43(1):10-12.
9. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci*. Mar 2004;59(3):255-263.
10. Theou O, Rockwood MR, Mitnitski A, Rockwood K. Disability and co-morbidity in relation to frailty: how much do they overlap? *Arch Gerontol Geriatr*. Sep-Oct 2012;55(2):e1-8.
11. Lutomski JE, Baars MA, Schalk BW, et al. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PloS One*. 2013;8(12):e81673.
12. van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol*. May 1998;51(5):367-375.
13. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc*. Jun 1992;40(6):585-588.
14. Mitnitski AB, Mogilner AJ, MacKnight C, Rockwood K. The mortality rate as a function of accumulated deficits in a frailty index. *Mech Ageing Dev*. 2002;123(11):1457-1460.
15. Lutomski JE, Baars MA, van Kempen JA, et al. Validation of a frailty index from the older persons and informal caregivers survey minimum data set. *J Am Geriatr Soc*. Sep 2013;61(9):1625-1627.
16. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8:24.
17. Cantril H. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press; 1965.
18. Hakkaart-van Royen L, Tan SS, Bouwmans CAM. Methoden en standaard kostprijzen voor economische evaluaties kostenonderzoek. College voor Zorgverzekeringen. 2010.
19. Hood K, Robling M, Ingledew D, et al. Mode of data elicitation, acquisition and response to surveys: a systematic review. *Health Technol Assess*. May 2012;16(27):1-162.
20. Ensrud KE, Ewing SK, Cawthon PM, et al. A comparison of frailty indexes for the prediction of falls, disability, fractures, and mortality in older men. *J Am Geriatr Soc*. Mar 2009;57(3):492-498.
21. Partridge JS, Harari D, Dhesi JK. Frailty in the older surgical patient: a review. *Age Ageing*. Mar 2012;41(2):142-147.
22. Rowe JW, Kahn RL. Successful aging. *Gerontologist*. Aug 1997;37(4):433-440.
23. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. Mar 2 2013;381(9868):752-762.
24. Gobbens RJ, van Assen MA, Luijkx KG, Wijnen-Sponselee MT, Schols JM. The Tilburg Frailty Indicator: psychometric properties. *J Am Med Dir Assoc*. Jun 2010;11(5):344-355.

25. Mitnitski AB, Graham JE, Mogilner AJ, Rockwood K. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatr.* Feb 27 2002;2:1.
26. Kulminski AM, Ukraintseva SV, Kulminskaya IV, Arbeev KG, Land K, Yashin AI. Cumulative deficits better characterize susceptibility to death in elderly people than phenotypic frailty: lessons from the Cardiovascular Health Study. *J Am Geriatr Soc.* May 2008;56(5):898-903.
27. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol.* Aug 2013;42(4):1012-1014.

Supplemental Table 1 Sociodemographic characteristics of participants who completed a written questionnaire by living situation

	Total (n=16,959)	Community-dwelling (n=16,322)	Residential care facility (n=637)
		Percentage	
Men	44.0	44.5	31.1
Education			
Primary school or lower	31.1	30.1	53.4
Secondary	47.9	48.4	36.9
Some college / College	21.0	21.5	9.7
Marital status			
Married / Cohabiting	59.0	60.0	33.8
Divorced / Single	10.7	10.5	15.6
Widowed	30.3	29.5	50.6
Nationality			
Native Dutch	91.0	90.9	94.6
First generation	4.3	4.3	2.9
Second generation	4.7	4.7	2.5
Self-reported health			
Excellent / Very good	14.3	14.7	5.4
Good	43.0	43.5	31.2
Fair / Poor	42.7	41.8	63.4
Self-reported health relative to previous year			
Much better/Somewhat better	8.8	8.9	5.5
Same	63.4	64.0	49.0
Somewhat worse/Much worse	27.8	27.0	45.5
Frailty	33.8	32.1	72.4
Multimorbidity	66.2	65.5	81.6
ADL limitation	26.6	25.1	64.5
		Mean (SD)	
Age (years)	76 (7)	76 (6)	84 (6)
Quality of life score	0.8 (0.2)	0.8 (0.2)	0.7 (0.2)
Psychological wellbeing	74.5 (17.7)	74.8 (17.6)	68.3 (19.4)

Note: Frailty was defined as a score ≥0.25 on 23-item frailty index derived from TOPICS-MDS. Multimorbidity was defined as the presence of ≥2 chronic conditions. Limitations in ADL were defined as requiring assistance with ≥1 item on the 6-item Katz Activities of Daily Living Index. Quality of life as measured using the EQ-5D summary score (score range -0.33 to 1.0, higher is better). Psychological wellbeing was measured using the Rand-36 Mini-Mental Health Scale (score range 0-100, higher is better).

Supplemental Table 2 Characteristics of participants who completed a written questionnaire by living situation

	Total (n=16,959)	Community-dwelling (n=16,322)	Residential care facility (n=637)
No complications (reference group)	26.6	27.4	9.4
Frailty only	2.2	2.2	2.1
Multimorbidity only	27.7	28.4	10.6
ADL limitations only	2.8	2.9	1.7
Multimorbidity and ADL limitations	8.6	8.7	5.7
Frailty and multimorbidity	13.0	13.0	12.8
Frailty and ADL limitations	1.4	1.3	3.8
Frailty, multimorbidity and ADL limitations	17.6	16.0	54.0
Total	**100%**	**100%**	**100%**

Note: Categories are mutually exclusive. Percentages are reported.

Supplemental Table 3 The association between self-reported quality of life scores and frailty, multimorbidity, and ADL limitations by living situation (based on written questionnaire data only)

	Community-dwelling	Residential care facility
No complications (reference group)	8.0 (7.8; 8.2)	8.3 (7.7; 8.9)
Frailty only	-1.2 (-1.3; -1.1)	-1.0 (-1.9; -0.03)
Multimorbidity only	-0.3 (-0.4; -0.3)	-0.2 (-0.8; 0.3)
ADL limitations only	-0.3 (-0.4; -0.2)	0.02 (-1.0; 1.0)
Multimorbidity and ADL limitations	-0.4 (-0.5; -0.3)	-0.5 (-1.2; 0.1)
Frailty and multimorbidity	-1.4 (-1.5; -1.3)	-1.4 (-1.9; -0.9)
Frailty and ADL limitations	-1.5 (-1.7; -1.3)	-1.4 (-2.1; -0.6)
Frailty, multimorbidity and ADL limitations	-1.5 (-1.6; -1.5)	-1.8 (-2.2; -1.3)

Note: The average differences (95% confidence interval) in quality of life scores are presented relative to participants reporting no complications. The model has been corrected for age, sex, education, and random effect by study project.

Supplemental Table 4 The association between care costs (€) and frailty, multimorbidity and ADL limitations by living situation (based on written questionnaire data only)

	Community-dwelling	Residential care facility
No complications (reference group)	1,595 (-1,533; 4,723)	35,704 (22,031; 49,377)
Frailty only	-1,118 (-2,337; 102)	-3,483 (-18,112; 11,145)
Multimorbidity only	-546 (-1,015; -77)	-1,425 (-9,446; 6,596)
ADL limitations only	-876 (-1,958; 206)	1,386 (-13,942; 16,714)
Multimorbidity and ADL limitations	-1,149 (-1,835; -463)	6,066 (-3,697; 15,828)
Frailty and multimorbidity	-560 (-1,170; 51)	8,099 (356; 15,841)
Frailty and ADL limitations	422 (-1,109; 1,953)	6,382 (-4,508; 17,273)
Frailty, multimorbidity and ADL limitations	449 (-147; 1,045)	1,473 (-5,266; 8,212)

Note: The average differences (95% confidence interval) in care costs are presented relative to participants reporting no complications. The model has been corrected for age, sex, education, and random effect by study project.

4.3

Advancing quality of life measures: Validating TOPICS Composite Endpoint

Submitted for publication as:

Hofman CS, **Lutomski JE**, Boter H, Buurman BM, de Craen AJM, Donders R, Olde Rikkert MGM, Makai P, Melis RJF. Examining the construct and known-group validity of a composite endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS).

Abstract

Background: Preference-weighted multi-faceted endpoints have the potential to facilitate comparative effectiveness research that incorporates patient preferences. The Older Persons and Informal Caregivers Survey – Composite endpoint (TOPICS-CEP) is potentially a valuable outcome measure for evaluating interventions in geriatric care as it combines multiple outcomes relevant to older persons in a single metric. The objective of this study was to validate TOPICS-CEP across different study settings (general population, primary care and hospital).

Methods: Data were extracted from TOPICS Minimum Dataset (MDS), a pooled public-access national database with information on older persons throughout the Netherlands. Data of 17,603 older persons were used. Meta-correlations were performed between TOPICS-CEP indexed scores, EuroQol5-D utility scores and Cantril's ladder life satisfaction scores. Mixed linear regression analyses were performed to compare TOPICS-CEP indexed scores between known groups, e.g. persons with versus without depression.

Results: In the complete sample and when stratified by study setting TOPICS-CEP and Cantril's ladder were moderately correlated, whereas TOPICS-CEP and EQ-5D were highly correlated. Higher mean TOPICS-CEP scores were found in persons who were: married, lived independently and had an education at university level. Moreover, higher mean TOPICS-CEP scores were found in persons without dementia, depression, and dizziness with falls. Similar results were found when stratified by subgroup.

Conclusion: This study supports that TOPICS-CEP is a robust measure which can be used in broad settings to identify the effect of intervention or of prevention in elderly care.

Introduction

Aging of the population has a major impact on the organization and delivery of healthcare. The shift from acute to chronic illnesses and the expected shortage of healthcare workers will be of particular importance.[1] To ensure high quality care for older persons, the evaluation and monitoring of three aspects of health care delivery need to be regularly evaluated: structure, process, and outcomes.[2] However, comparing outcomes in older persons is challenging. Firstly, the health states of older persons are complex, as older individuals often present different combinations of chronic multi-morbidity and functional limitations.[3] Secondly, interventions often influence a broad range of health domains both directly and indirectly. For example, occupational therapy aims to enable people who have physical restrictions to achieve greater independence. By engaging in meaningful social activities, health and psychological wellbeing are also indirectly and positively influenced.[4] Thus, occupational therapy can improve both physical and mental wellbeing. The two obstacles can be circumvented if the important outcome parameters are collected and combined into a preference-weighted composite endpoint (CEP).[5,6]

In 2008, the Dutch Care for the Elderly Programme was commissioned by the Ministry of Health, Welfare and Sport with the guiding principles of improving care, quality of life, and self-management among older persons. As part of this Programme, The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS) was developed to uniform collection of outcome measures.[7] To promote comparability between research studies, a preference-weighted CEP was established for TOPICS-MDS based on the health state valuations of older persons and informal caregivers. This CEP (referred to as TOPICS-CEP) was designed as a multi-faceted outcome measure applying weights derived from older persons' priorities for different outcomes to assist in the evaluation of interventions in older persons.[8]

TOPICS-CEP has been previously developed using a vignette study in which 200 persons participated. Profiles of older persons (vignettes) were used to obtain the preference weights for TOPICS-CEP's components.[8] The aim of this current study was to determine TOPICS-CEP's convergent and known-groups validity in large heterogeneous samples of older persons aged 65 years and older and across general population, primary care and hospital setting.

Methods

Data source

Data were derived from TOPICS-MDS (www.topics-mds.eu), which is a public data repository designed to capture essential information on the physical and mental wellbeing of older persons and informal caregivers in the Netherlands. A detailed description of TOPICS-MDS has been presented elsewhere.[7] Briefly, TOPICS-MDS consists of pooled data from various research projects which differ across study design, sampling framework, and inclusion criteria. All data were cleaned locally using a standardized protocol. Anonymized individual-level data were then submitted to a central institution (Radboud University Medical Center, Nijmegen, the Netherlands) for further validation checks and creation of the pooled dataset. Since various research projects submit information to TOPICS-MDS, the database is dynamic in nature and thus regularly updated with new observations.

Our present analysis uses the first version of the dataset available as of January 2013 and is based on 41 studies with data available on 32,310 older persons. Studies which omitted TOPICS-CEP data points by design were excluded from this study. This, resulted in a final study sample of 17,603 older persons.

TOPICS-MDS is a fully anonymized dataset available for public access, and therefore this analysis was exempt from ethical review (Radboud University Medical Center Ethical Committee review reference number: CMO: 2012/120).

Measures

TOPICS-CEP

TOPICS-CEP score is a preference-weighted index ranging from 0 (worst possible state) to 10 (best possible state) that combines 42 data points representing eight domains: morbidities (list of 17 pre-defined conditions widely used in the Netherlands),[9] functional limitations (Katz index of independence),[10] emotional wellbeing (mental health subscale of the RAND-36),[11] pain experience (pain dimension of the EQ-5D),[12] cognitive problems (cognition dimension of the EQ-5D+C),[12] social functioning (item 10 from the RAND-36),[11] self-perceived health (item from the RAND-36)[11] and self-perceived quality of life (phrasing similar to self-perceived health item from the RAND-36).[11] The components vary in scale range and preference weight. More detailed information about TOPICS-CEP, including a description of the data points, can be found elsewhere.[13] Briefly, TOPICS-CEP score is calculated in four steps. Firstly, data points are coded in the same direction by means of reversed scoring. Secondly, all items that belong to the same health domain are aggregated into one component. Thus, 17 morbidity items are combined into the component *number of morbidities*,

15 items regarding functional limitations into *number of functional limitations*, and 5 emotional well-being items into *raw emotional well-being score*. Thirdly, a raw TOPICS-CEP score is calculated by means of applying the preference weights for the Dutch population aged 65 years and over.

> *Raw TOPICS-CEP score = 9.00 (Intercept) – [0.13 x morbidities] – [0.12 x functional limitations] – [0.03 x emotional wellbeing] – [0.03 x pain experience] – [0.14 x cognitive problems] – [0.01 x social functioning] – [0.17 x self-perceived health] – [0.02 x self-perceived quality of life].*

Finally, the raw TOPICS-CEP score is transformed into an indexed score (referred to as TOPICS-CEP score) ranging 0 to10.

> *TOPICS-CEP score = [(raw TOPICS-CEP score – minimum raw TOPICS-CEP score)/raw score range] x 10. = [(raw TOPICS-CEP score – 2.58)/5.90] x 10.*

In this current study, only missing data points were allowed for the aggregated TOPICS-CEP components morbidities, functional limitations and emotional wellbeing. The thresholds used were less than five missing values for morbidities and functional limitations respectively, and less than two missing values for emotional wellbeing. Estimation for these data points was done by pro-rating the score. For instance, the component functional limitations includes 15 items and the scale range is zero to 15; when 12 items are answered and the sum of the answered items is six, then score pro-rating =[(6/12) x 15]=7.5.

Other measures

The Cantril's life satisfaction score is a one-dimensional index ranging from zero (completely unsatisfied with life) to ten (completely satisfied with life) and measures self-perceived general QOL.[14] We used a modified version of Cantril's self-anchoring ladder where respondents were asked to rate their present life on a scale between zero and ten.

The EuroQol-5D (EQ-5D) utility score measures health related QOL (HRQOL).[15] Five dimensions (mobility, self-care, daily activities, pain and discomfort, anxiety and depression) with three levels each (1=no problems, 2=moderate problems, and 3=extreme problems) are combined into one utility score by means of applying the scoring values for the Dutch population.[15] The EQ-5D utility score ranges from -0.33 to 1.00 where a score of less than zero is indicative of a health state worse than death.[15]

Socio-demographic characteristics included in our analyses were marital status, living arrangements, and education level. Included clinical data points were dementia, depression, and dizziness with falls.

Convergent validity

Convergent validity refers to how closely a measure is related to other measure of the same construct**.** We examined convergent validity of TOPICS-CEP score with the Cantril's life satisfaction score and the EQ-5D utility score respectively.[14,15] Convergent validity is determined by the correlation between the outcome measures.

Hypotheses

We anticipated a moderate positive correlation between TOPICS-CEP score and the Cantril's life satisfaction score, because TOPICS-CEP intends to measure a broader concept than self-perceived general QOL. In contrast, we expected a strong positive correlation between TOPICS-CEP score and the EQ-5D utility score as both measures combines multiple outcomes, however they do have a different score range [TOPICS-CEP: 0-10 versus EQ-5D: -0.33 – 1.0].

Known-group validity

After examining the convergent validity, we examined whether groups with different marital status, living arrangements, education levels and the presence or absence of the chronic conditions dementia, depression, and dizziness with falls could be distinguished based on their TOPICS-CEP scores. Thus, we assessed whether baseline TOPICS-CEP scores were significantly different between groups.

Hypotheses

We expected higher scores in persons who are married or cohabiting compared to widowers and in those who live with others (e.g. partner or children) compared to those who live alone because long lasting relationships positively influences (mental) health status.[16] Similarly, we expected to find higher scores in older persons living independently compared to those living in an institutionalized facility. This is largely due to institutionalized older persons often require more assistance with daily activities and thus may fear their loss of independence, control and dignity.[17] Furthermore, we anticipated to find lower scores in subgroups of persons with dementia, depression, or dizziness with falls than in persons without these conditions. Such conditions have wide-reaching effects and would likely negatively impact other domains included in TOPICS-CEP.[18-20]

Generalizability

To examine whether the validation results for TOPICS-CEP are generalizable across different settings, we performed additional analyses using the complete study sample as well as stratified across three major study settings: older persons in primary care setting, general older population, and hospitalized older persons.

Analyses

Feasibility was assessed by calculating the number of missing values for TOPICS-CEP. Floor and ceiling effects were assessed by reporting the proportion of respondents with minimum and maximum TOPICS-CEP scores, respectively. A floor or ceiling effect of 15% was considered the maximum acceptable.[21]

Since TOPICS-MDS is a pooled dataset, we applied meta-analytical techniques to account for clustering within individual research projects. Pearson's correlations were used to examine convergent validity between TOPICS-CEP, Cantril's life satisfaction scale, and EQ-5D utility score within each study. To calculate the pooled correlation coefficients random effects meta-correlations were performed.[22] Correlations below 0.3 were referred to as weak, between 0.3 and 0.5 as moderate, and above 0.5 as strong.[23]

Known group validity was examined by determining significant differences in mean TOPICS-CEP index scores. Mixed linear regression analyses were used to compare the scores between groups and to examine whether differences between groups were still present when adjusted for age and gender. To account for clustering within individual research projects the models included random intercepts for project. The models were constructed based on *a priori* expectations. Differences between parameter estimates smaller than 15% were considered to be acceptable. Analyses were performed using SPSS version 20.0 (SPSS IBM, New York, USA) and the Meta package in R (Foundation for statistical computing, Vienna, Austria).[22]

Results

Sample characteristics

Data from 17,603 older persons from 28 projects were included in this study. The majority of the study sample were women (N= 10,817, 61.5%) and the mean (±SD) age was 79 (7) years. Overall, the sample consisted of 7,849 (44.9%) subjects living independently with others, 8,187 (46.7%) were married or cohabiting, and 7,965 (46.7%) had a secondary education level. The conditions dementia, depression, and dizziness with falls were present in; 962 (5.6%), 1,558 (9.1%), and 2,495 (14.6%)

subjects of the study sample respectively. The socio-demographic distribution within the subgroups (primary care (N=11,892), general population (N=3,331), and hospital (N=1,534)) were similar to the combined sample.

Outcomes

Of the 17,603 participants, the majority had no missing data points for TOPICS-CEP: 88.7% (N=15,612), Cantril's ladder: 91.9% (N=16,178) and EQ-5D: 96.6% (N=17,006). The means (±SD; minimum and maximum scores achieved) were TOPICS-CEP: 7.37 (1.23; 1.88 – 10.0); Cantril's ladder: 7.12 (1.40;0.0 – 10.0); and EQ-5D: 0.63 (0.29; -0.33 – 1.0). Table 1 gives an overview of the mean (±SD) scores and floor and ceiling effects for the complete sample and stratified by subgroup. The highest values possible for TOPICS-CEP, Cantril's ladder, and EQ-5D was reported for 18 (0.1%), 379 (2.2%), and 2,009(11.4%) older persons respectively. For each outcome measure, the lowest value possible was calculated for less than 1% of the subjects.

Table 1 The mean (±SD) scores and floor and ceiling effects for the complete sample and stratified by subgroup

	Mean (SD)	Floor N (%)	Ceiling N (%)
Complete study sample (N=17,603)			
TOPICS-CEP	7.37 (1.23)	0 (0.0)	18 (0.1)
Cantril's ladder	7.12 (1.40)	22 (0.1)	379 (2.2)
EQ-5D	0.63 (0.29)	0 (0.0)	2,009 (11.4)
Subgroups by study setting			
Primary care setting (N=11,892)			
TOPICS-CEP	7.44 (1.15)	0 (0.0)	12 (0.1)
Cantril's ladder	7.11 (1.42)	17 (0.1)	257 (2.2)
EQ-5D	0.61 (0.28)	0 (0.0)	1,100 (9.2)
General population (N=3,331)			
TOPICS-CEP	7.37 (1.40)	0 (0.0)	5 (0.2)
Cantril's ladder	7.07 (1.34)	4 (0.1)	47 (1.4)
EQ-5D	0.72 (0.26)	0 (0.0)	653 (19.6)
Hospital (N=1,534)			
TOPICS-CEP	7.48 (1.20)	0 (0.0)	1 (0.1)
Cantril's ladder	7.36 (1.35)	1 (0.1)	49 (3.2)
EQ-5D	0.61 (0.30)	0 (0.0)	210 (13.7)

When stratified by subgroup the mean (±SD) scores showed similar patterns. For each outcome measure the lowest value possible was achieved by less than 1% of the older persons whereas the highest possible value for EQ-5D was calculated for 19.6% (N=653) and 13.7% (N=210) of the older persons sampled from the general population and hospital respectively.

Convergent validity

Table 2 gives an overview of the meta-correlation coefficients and the 95% CI. Expectedly, TOPICS-CEP and Cantril's ladder were moderately correlated in the overall sample and subgroups Complete sample: r=0.43; Primary care: r=0.41; General population: r=0.50; Hospital: r=0.43. In comparison, TOPICS-CEP and the EQ-5D were highly correlated [Complete sample: r=0.63; Primary care: r=0.60; General population: r=0.71; Hospital: r=0.57].

Known group validity

Table 3 illustrates the association between TOPICS-CEP scores and sample characteristics. In line with our expectations, higher mean TOPICS-CEP scores

Table 2 Meta-correlation coefficients and the 95% CI of the outcome measures TOPICS-CEP, Cantril's ladder, and EQ-5D utility score for the complete study sample and stratified by subgroup

	TOPICS-CEP	Cantril's
	r (95% CI)	r (95% CI)
Complete study sample (N=17,603)		
Cantril's ladder	0.43 (0.39-0.48)	
EQ-5D	0.63 (0.58-0.67)	0.34 (0.28-0.40)
Subgroups by study setting		
Primary care (N=11,892)		
Cantril's ladder	0.41 (0.33-0.48)	
EQ-5D	0.60 (0.52-0.67)	0.31 (0.21-0.41)
General Population (N=2,221)		
Cantril's ladder	0.53 (0.51-0.56)	
EQ-5D	0.71 (0.68-0.74)	0.43 (0.35-0.50)
Hospital (N=1,534)		
Cantril's ladder	0.43 (0.35-0.51)	
EQ-5D	0.57 (0.51-0.62)	0.29 (0.25-0.34)

Table 3 Associations between TOPICS-CEP scores and sample characteristics for the complete study sample and stratified by subgroup

	Overall Unadjusted	Overall Adjusted	Primary care	General Population	Hospital
	Estimates (95% CI)	Estimates (95% CI)	Estimates (95% CI)	Estimates (95% CI)	Estimates (95% CI)
Age					
78 years old (Reference)	7.27 (7.05, 7.49)	--	7.29 (6.99, 7.57)	7.50 (6.95, 9.04)	7.08 (5.66, 8.50)
Per additional year	0.03 (0.02, 0.04)	--	0.04 (0.04, 0.04)	0.06 (0.05, 0.07)	0.04 (0.03, 0.05)
Gender					
Male (Reference)	7.56 (7.32, 7.80)	--	7.51 (7.21, 7.82)	7.99 (7.45, 8.53)	7.47 (6.08, 8.86)
Female	-0.39 (-0.43, -0.36)	--	-0.34 (-0.38, -0.30)	-0.48 (-0.58, -0.38)	-0.55 (-0.67, -0.43)
Marital status					
Married/Cohabiting (Reference)	7.50 (7.26-7.74)	7.19 (6.97, 7.42)	7.45 (7.14, 7.76)	7.96 (7.43, 8.48)	7.42 (6.15, 8.70)
Widowed	-0.37 (-0.41, -0.34)	-0.08 (-0.13, -0.04)	-0.29 (-0.33, 0.24)	-0.62 (-0.73, -0.52)	-0.53 (-0.66, -0.40)
Other	-0.22 (-0.28, -0.16)	-0.12 (-0.18, -0.06)	-0.17 (-0.24, -0.10)	-0.34 (-0.51, -0.19)	-0.40 (-0.59, -0.20)
Living arrangement					
Independent, alone (Reference)	7.37 (7.14, 7.59)	7.27 (7.05, 7.49)	7.32 (7.01, 7.64)	7.75 (7.32, 8.17)	7.01 (5.63, 8.40)
Independent, with others	0.19 (0.16, 0.23)	0.01 (-0.03, 0.05)	0.16 (0.12, 0.20)	0.21 (0.10, 0.32)	0.40 (0.28, 0.52)
Dependent	-1.01 (-1.08, -0.95)	-0.86 (-0.94, -0.82)	-0.78 (-0.87, -0.70)	-1.32 (-1.44, -1.19)	-1.03 (-1.29, -0.78)
Educational background					
Primary or less (Reference)	7.13 (6.88, 7.37)	7.02 (6.80, 7.25)	7.15 (6.84, 7.46)	7.32 (6.79, 7.88)	6.94 (5.59, 8.30)
Secondary/technical training	0.27 (0.23, 0.31)	0.17 (0.13, 0.21)	0.23 (0.18, 0.27)	0.45 (0.34, 0.55)	0.29 (0.15, 0.43)
College/some college	0.44 (0.39, 0.50)	0.30 (0.24, 0.35)	0.38 (0.32, 0.45)	0.76 (0.61, 0.92)	0.43 (0.26, 0.61)

Dementia					
No (Reference)	7.43 (7.10, 7.77)	7.23 (6.93, 7.54)	7.42 (7.00, 7.83)	7.88 (7.20, 8.56)	7.18 (5.46, 8.89)
Yes	-1.13 (-1.21, -1.04)	-1.04 (-1.13, -0.95)	-0.86 (-0.96, -0.75)	-1.62 (-1.80, -1.44)	-1.75 (-2.41, -1.08)
Depression					
No (Reference)	7.42 (7.15, 7.71)	7.25 (6.99, 7.51)	7.44 (7.09, 7.78)	7.78 (7.11, 8.47)	7.25 (5.80, 8.69)
Yes	-1.16 (-1.21, -1.10)	-1.16 (-1.22, -1.11)	-1.11 (-1.18, -1.05)	-1.44 (-1.61, -1.27)	-1.28 (-1.51, -1.06)
Dizziness with falls					
No (Reference)	7.49 (7.20, 7.75)	7.29 (7.04, 7.54)	7.47 (7.13, 7.81)	7.85 (7.22, 8.49)	7.33 (5.84, 8.82)
Yes	-1.07 (-1.11, -1.01)	-1.01 (-1.05, -0.96)	-1.05 (-1.10, -0.99)	-1.18 (-1.32, -1.05)	-1.15 (-1.31, -0.99)

were found in older adults who were married, lived independently and had a higher education level, respectively. Moreover, the mean TOPICS-CEP scores were higher in the persons without dementia, depression and dizziness with falls, respectively. Table 4 illustrates the relationships between TOPICS-CEP scores and sample characteristics adjusted for gender and age. The parameter estimates of marital status and education level remained significant (p-values <0.05) after adjustments; however, these exceeded the 15% threshold of change. Thus, for example the average difference between TOPICS-CEP scores of persons who were married or cohabiting versus those who had a deceased partner was still significantly different; however the difference between the scores decreased from 0.37 to 0.08. Furthermore, the parameter estimate of living independently with others was no longer significant after adjustment for gender and age. Without adjustment, the average difference TOPICS-CEP scores of persons living independently alone versus living independently with others were 0.19 points and with the adjustment the difference was 0.01 point. When stratified by subgroup similar results were found (data not shown).

Discussion

The purpose of this study was to determine convergent and known group validity of TOPICS-CEP in a large and heterogeneous sample of persons aged 65 years and older. Preference-weighted composite endpoints such as TOPICS-CEP have the potential to facilitate comparative effectiveness research, thus it is important to establish the validity of these kinds of endpoints prior to their use in the population of interest.

In this current study, TOPICS-CEP was able to accurately represent the heterogeneous composition of the overall study population. TOPICS-CEP scores obtained covered most of the entire current score range of the index and there were no floor or ceiling effects found in the total sample nor in the subsample taken from general population, primary care or hospital settings. This is important for its performance as an outcome measure. At the same time, the EQ-5D utility scores showed considerably larger ceiling effects in the general population sample. The most plausible reason why this specific subgroup exhibited this effect would be that the persons from the general population sample were less frail compared to those from the primary care and hospital sample.

Our correlation analyses revealed significant associations between TOPICS-CEP score versus Cantril's ladder and EQ-5D utility score. The stronger correlation between TOPICS-CEP and EQ-5D indicates that the TOPICS-CEP measures important aspects of health. As expected, the correlation between TOPICS-CEP

and Cantril's ladder was moderate because the two outcomes measure different concepts. Moreover, our findings supported our hypothesis that there would be a strong correlation between TOPICS-CEP components and the EQ-5D dimensions.

TOPICS-CEP scores adhered expected patterns across marital status, living arrangements, and education level. Additionally, TOPICS-CEP was able to distinguish subjects who had dementia, depression, and dizziness with falls even when adjusted for age and gender. These findings further support the overall validity of the tool.

Our results indicate that there were no floor or ceiling effects for TOPICS-CEP in the different settings. However, similar to other studies we found a ceiling effect for the EQ-5D utility score as the percentage of persons with the highest possible EQ-5D utility score of 1.00 exceeded the 15% threshold. These ceiling effects may be due to a small range of responses (3 levels per item).[24,25]

There are several limitations to consider. Even though a large heterogeneous sample was used to validate TOPICS-CEP, institutionalized older persons were underrepresented in our sample. Secondly, additional research is required to examine other important properties of TOPICS-CEP, such as minimal clinically important difference and the sensitivity to detect change. For these reasons, longitudinal validation would be beneficial.

In conclusion, preference-weighted multi-faceted endpoints have the potential to facilitate comparative effectiveness research that incorporates patient preferences. This study supports that TOPICS-CEP is a good option for researchers who need an outcome measure to assess important outcomes for older persons even when it is across a range of differently functioning subpopulations. TOPICS-CEP is a robust measure which can be used in broad settings to identify the effect of intervention or of prevention in elderly care. It deserves further spread as the various outcome domains included in the measure are of great importance to the older population.

References

1. Kovner CT, Mezey M, Harrington C. Who cares for older adults? Workforce implications of an aging society. *Health Aff (Millwood).* 2002;21(5):78-89.
2. Donabedian A. Evaluating the quality of medical care. *Milbank Q.* 1966;44(3):Suppl:166-206.
3. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci.* 2004;59(3):255-263.
4. Park NS. The relationship of social engagement to psychological well-being of older adults in assisted living facilities. *J Applied Gerontol.* 2009;28(4):461-481.
5. Erickson P, Kendall EA, Anderson JP, Kaplan RM. Using Composite Health-Status Measures to Assess the Nations Health. *Med Care.* 1989;27(3):S66-S76.
6. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care.* 1989;27(3 Suppl):S217-232.
7. Lutomski JE, Baars MA, Schalk BW, et al. The Development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): A Large-Scale Data Sharing Initiative. *PloS One.* 2013;8(12):e81673.
8. Hofman CS, Makai P, Boter H, et al. Establishing a composite endpoint for measuring the effectiveness of geriatric interventions based on older persons' and informal caregivers' preference weights: a vignette study. *BMC Geriatr.* 2014;14(1):51.
9. Lokale en nationale monitor gezondheid (In Dutch only). https://www.monitorgezondheid.nl/gezondheidindicatoren.aspx. Accessed May 15, 2013.
10. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. *J Am Geriatr Soc.* 1992;40(6):585-588.
11. VanderZee KI, Sanderman R, Heyink JW, de Haes H. Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. *Int J Behav Med.* 1996;3(2):104-122.
12. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *J Clin Epidemiol.* 1999;52(4):293-301.
13. Hofman CS, Melis RJF, Donders ART, Makai P, OldeRikkert MGM. The Older Persons and Informal Caregivers Survey - Composite Endpoint (TOPICS-CEP) Guidelines, First version http://topics-mds.eu. Accessed 1 April, 2014.
14. Cantril H. *The Pattern of Human Concerns.* New Brunswick, N.J.: Rutgers University Press; 1965.
15. Szende A, Oppe M, Devlin NJ. *EQ-5D Value Sets : Inventory, comparative review and user guide.* Dordrecht ; [London]: Springer; 2007.
16. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull.* 1995;117(3):497-529.
17. Woolhead G, Calnan M, Dieppe P, Tadd W. Dignity in older age: what do older people in the United Kingdom think? *Age Ageing.* 2004;33(2):165-170.
18. Crooks VC, Lubben J, Petitti DB, Little D, Chiu V. Social network, cognitive function, and dementia incidence among elderly women. *Am J Public Health.* 2008;98(7):1221-1227.
19. Faes MC, Reelick MF, Joosten-Weyn Banningh LW, Gier M, Esselink RA, Olde Rikkert MG. Qualitative study on the impact of falling in frail older persons and family caregivers: foundations for an intervention to prevent falls. *Aging Ment Health.* 2010;14(7):834-842.
20. Malhotra R, Chei CL, Ostbye T, Chan A, Matchar DB. Older person behavioral and psychological symptoms (BPS) and functional limitations mediate the association between older person cognitive impairment and depressive symptoms in the caregiver. *Arch Gerontol Geriatr.* 2014;58(2):269-277.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidem.* 2007;60(1):34-42.
22. Schwarzer G. Meta: Meta-Analysis with R. 2010; http://CRAN.R-project.org/package=meta. Accessed March 6th, 2014.

23. Cohen J. Set Correlation and Contingency-Tables. *Appl Psych Meas*. 1988;12(4):425-434.
24. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873-884.
25. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22(7):1717-1727.

Adapted in part from:

van den Brink D, **Lutomski JE**, Qin L, den Elzen WP, Kempen GI, Krabbe PF, Steyerberg EW, Muntinga M, Moll van Charante EP, Bleijenberg N, Olde Rikkert MG, Melis RJ, on behalf of TOPICS-MDS Consortium. TOPICS-MDS: a versatile resource for generating scientific and social knowledge for elderly care. *Tijdschrift voor Gerontologie en Geriatrie*. 2015 Apr;46(2):78-91. doi: 10.1007/s12439-015-0127-3.

5

Conclusions & General Discussion

Key findings

- The 15-item ADL/IADL scale demonstrates poor discriminative ability in higher functioning older persons. In future versions of the MDS, the physical functioning domain should be revised to better represent this segment of the population.
- Survey mode influences reporting of urinary incontinence, walking, and meal preparation in residential care settings.
- Whereas a ceiling effect for the EQ-5D, an instrument measuring health-related quality of life, is observed in the overall study population, such is not the case within subgroups presenting with major geriatric conditions.
- The validity of the CarerQol, an instrument measuring care-related quality of life, is maintained irrespective of sampling framework (general population, primary care centers, or hospital settings) though differential reporting is observed by survey mode.
- A frailty index derived from TOPICS-MDS provides a valid indication of frailty.
- When evaluating quality of life and health care costs, frailty is a useful independent covariate in addition to multimorbidity and ADL limitations.
- A preference-weighted composite endpoint derived from TOPICS-MDS provides a valid indication of health-related quality of life useful for evaluation studies.

5.1. Summary of findings

Although data sharing has garnered theoretical support throughout the scientific community, it has not been widely carried out in practice. This is due, in part, to a lack of collaboration among researchers. The Older Persons and Informal Caregivers Survey – Minimum Dataset (TOPICS-MDS) thus represents a pioneering achievement since its development not only required the commitment of the eight Dutch university medical centers but also numerous researchers across the Netherlands. All principal investigators receiving funding from the National Care for the Elderly Programme agreed to incorporate TOPICS-MDS into their study protocols. The product was a centrally pooled, public access database with standardized health outcomes collected from over 60 independent research projects. Undoubtedly, the development of the database represents a major milestone in data sharing and confers many research opportunities. However, any new data collection tool should be approached with circumspection until the accuracy and quality of the data are appropriately evaluated.

Thus, the overarching aim of this thesis was to validate the data compiled in TOPICS-MDS. To meet this aim, Chapter 3 reviewed the measurement properties of the three primary instruments included in TOPICS-MDS: The Katz Index of Independence in Activities of Daily Living, the EQ-5D, and the CarerQol. Chapter 4 evaluated two derived variables for measuring frailty and health-related quality of life. Since frailty is a key indicator in geriatric research, Section 4.1 validated the derivation of frailty index (i.e. a composite measure based on a spectrum of health problems) using TOPICS-MDS data, and Section 4.2 examined the added benefit of measuring frailty in relation to multimorbidity and activities of daily living. Given the growing interest in more comprehensive indicators for wellbeing, Section 4.3 validated a preference-weighted composite endpoint for health-related quality of life (referred to as TOPICS-CEP).

Chapter 3, Sections 3.1 and 3.2 focused on the modified Katz Index of Independence in Activities of Daily Living, which assessed whether older persons required assistance with 15 basic and instrumental activities of daily living (ADL and IADL respectively). In Section 3.1, a Rasch analysis was applied to determine whether the difficulty level of these 15 items were rated similarly among older persons residing in the community versus residential care facilities as well as if the scale possessed robust discriminative ability. This analysis found minor differences in rating of item difficulty, thus inhibiting direct comparative analyses. If such analyses are undertaken, item weighting may be necessary to standardize item difficulty levels between these two settings. Moreover, the scale poorly discriminated between higher functioning community-dwelling

older persons. In future versions of TOPICS-MDS, the physical functioning domain should be revised to better represent this segment of the population. In a complementary analysis presented in Section 3.2, a Rasch analysis was applied to examine whether survey mode (i.e. written questionnaire versus interview) resulted in differential reporting of activities in residential care settings. Survey mode was found to influence reporting of urinary incontinence, walking, and meal preparation, suggesting that either weighting or interaction terms may be required when analyzing pooled, mixed-mode data.

Chapter 3, Section 3.3 reviewed the measurement properties of the EQ-5D, an instrument measuring health-related quality of life, across four major geriatric conditions: hearing disorders, joint damage, urinary incontinence, and dizziness with falls. Outcomes were compared to the overall study population as well as a 'healthy' subgroup reporting no major chronic conditions. Irrespective of the geriatric condition subgroup, individual domains of the EQ-5D correlated well with other survey items measuring similar concepts and adhered to *a priori* hypotheses regarding sociodemographic associations, supporting construct validity. Ceiling effects were observed in the overall study population and the 'healthy' subgroup but not within the geriatric condition subgroups. Thus, although the pooled dataset for TOPICS-MDS initially suggests poor discriminative ability for the EQ-5D, such is not the case within subgroups presenting with major geriatric conditions.

Chapter 3, Section 3.4 validated the Care-Related Quality of Life Instrument (CarerQol) across two different study design features, sampling framework (general population versus different care settings) and survey mode (interview versus written questionnaire). The observed correlations between the CarerQol domains and a generic measure of happiness and self-reported burden supported the instrument's construct validity. Based on multivariable, multivariate mixed models, sampling framework had only a minor influence on reporting levels; however, survey mode seemed to modestly influence self-reported burden. Future analyses should account for the impact of survey mode on care-related quality of life reporting; the use of an interaction term in multivariable modelling represents one option to address this issue.

Chapter 4, Sections 4.1 and 4.2 examined the validity and operationalization of the frailty index. In Section 4.1, a long (46-item) and short (23-item) frailty index derived from TOPICS-MDS was correlated with a frailty index based on clinical assessment. The long and short TOPICS frailty indices were moderately correlated with the independent, clinically assessed frailty index ('reference standard'). Furthermore, both indices met *a priori* hypotheses regarding age and sex distributions, thus supporting their validity. Section 4.2 addressed the operationalization of frailty and its relation to multimorbidity and ADL limitations.

Frailty, multimorbidity, and ADL limitations were reported for a substantial proportion of older persons sampled in TOPICS-MDS, and older persons experiencing problems across the three domains reported the poorest quality of life scores and the highest health care costs. Based on multivariable mixed models measuring quality of life scores and health care costs, frailty was found to be an important independent covariate in addition to multimorbidity and ADL limitations.

Chapter 4, Section 4.3 validated a preference-weighted Composite End Point (CEP) for health-related quality of life derived from TOPICS-MDS (referred to as TOPICS-CEP). Based on meta-correlations, TOPICS-CEP was moderately correlated with Cantril's Self-Anchoring Ladder (i.e. an overall quality of life measure) and highly correlated with the EQ-5D (i.e. a health-related quality of life measure). Based on multivariable mixed models, TOPICS-CEP scores were higher in persons who were married, lived independently, or university educated whereas lower in older persons with dementia, depression, or dizziness with falls. The TOPICS-CEP behaved in accordance with *a priori* hypotheses, supporting construct and known-group validity. This study supported the use of TOPICS-CEP as a meaningful indicator of health-related quality of life in evaluation research.

5.2. General discussion

From a minimum dataset towards an integrated care database

The current health care model emerged from the need to treat acute or singular chronic conditions in younger populations. However, as modern society ages, health care providers and public health professionals are now working in an environment incongruous with the needs of older persons with increasingly complex health profiles.[1,2] Their symptoms often reflect an interaction between chronic multimorbidity[2] and age-associated frailty[3] which threaten or actually lower quality of life. For these individuals, the disjointed treatment and monitoring of single morbidities without regard for overall wellbeing does not promote healthy aging.[1] Quite the contrary, this narrow focus stifles concomitant gains in physical, mental, and social health that are often witnessed when a more integrated approach to health care is taken. Yet, integrated care represents more than an intervention to facilitate discourse between different health care providers attending to the same patient. In its broadest sense, integrated care represents "a step in the process of health systems and health care delivery becoming more complete and comprehensive"[4], thereby shifting health and health care conceptualization towards a more nuanced interpretation. While

there has been growing consensus that integrated care is the new face of medicine, achieving this model in practice is less straightforward. Transitioning from the current health care model to an integrated care model necessitates the re-configuration of large, established health care systems on both the individual and organizational levels.[4,5] Given the enormity of this challenge, the European Innovation Partnership on Active and Healthy Ageing (European 2020 Initiative) have endorsed a European-based integrated care action group. This group was established to promote enhanced case management of chronic conditions, the use of innovative tools and services to scale-up integrated care programs, and the sharing of knowledge between European partners.[6] This campaign has further acknowledged the importance of local initiatives to expand this more comprehensive approach to care.

With this backdrop, The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS) was created. Guided by the ethos of integrated care, TOPICS-MDS was designed to not only record relevant morbidity outcomes, but also provide insight into physical, mental, and social wellbeing of older persons and their caregivers. During its development phase, brevity was also viewed as a crucial aspect so that the survey could be easily incorporated into a range of academic and governmental settings. The product was a highly accessible data collection tool with a broad scope to assist health care providers, researchers, public health practitioners, and policy makers in better evaluating overall wellbeing, quality of care, and the effectiveness of interventions. In this way, TOPICS-MDS represents a small but critical step in achieving a more comprehensive interpretation of health and wellbeing and evaluating 21st century integrated health care.

What TOPICS-MDS offers

TOPICS-MDS is relatively short data collection tool with only 51 data points for older persons and 27 data points for caregivers. In brief, for older persons, information is collected on demographics, morbidity, quality of life, functional limitations, mental health, social functioning, and health service utilisation. For informal caregivers, information is collected on demographics, hours of informal care, and quality of life. Yet, the richness of the dataset not only lies in the existing elements but also the additional variables which can be derived from the dataset.

For instance, frailty is a central concept in geriatric research and represents a heightened state of physical, mental and social vulnerability.[3] Given its association with increased risk of functional dependency[7], falls[8], institutionalization[9], and death[10], frailty serves as an important case-mix variable for clinical case management and resource allocation in older populations. Using TOPICS-MDS,

a frailty index can be derived by calculating the proportion of health problems observed in an individual versus the number of health problems screened in the survey.[11] The frailty index has been shown to be a robust indicator for vulnerability and morality in other settings[10], and this thesis has not only validated its use in TOPICS-MDS but has also shown that it is a valuable indicator for quality of life and health care costs.

Moreover, outcomes collected in TOPICS-MDS may be analyzed separately (i.e. per item or item subset) or combined into a preference-weighted Composite End Point (TOPICS-CEP). TOPICS-CEP was developed to provide users of the database with the option of a multi-dimensional outcome measure which assesses eight domains (morbidity, limitations in activity, emotional well-being, pain experience, cognitive functioning, social functioning, self-perceived health, and self-perceived quality of life).[12] To account for relative differences in the perceived importance of these domains, preference weights were established by asking 124 older persons and 76 caregivers to rate domains using vignettes.[12] As shown in this thesis, TOPICS-CEP is a valid measure for measuring health-related quality of life.

TOPICS-MDS as a stimulus for data sharing and waste reduction

The vision of TOPICS-MDS was to be more than a tool to facilitate integrated care but also a broader initiative to promote data sharing. Developed as part of the National Care for the Elderly Programme (NPO), information on all older persons and caregivers who participated in NPO-funded projects was centrally collated to create a national, public access data repository. This project required the cooperation of numerous researchers through the Netherlands working on more than 60 different research studies. A large data sharing initiative like TOPICS-MDS is highly relevant in the current research climate. There has been a philosophical transition in scientific and political arenas as greater emphasis is being placed on standardization of health outcomes and open access data. In a recent letter (October 2013) to the Chairman of the House of Representatives in the Netherlands, Minister Edith Schippers (Minister of Health, Welfare, and Sport) posited that standardization and accessibility of health data are essential elements in creating a sustainable information system with timely and reliable data.[13] Moreover, in light of the decentralization of the public health sector to smaller localities under the Social Support Act (WMO 2015), standardization is critical to ensure regional and longitudinal comparisons can continue to be drawn across the Netherlands.

Funding bodies have also begun to include data accessibility to external third parties as part of their framework. The three largest research funding bodies in the Netherlands, The Netherlands Organisation for Health Research

and Development (ZonMw), The Netherlands Organisation for Scientific Research (NWO), and The Netherlands Academy of Science (KNAW), have all taken initial steps to ensure that well-annotated, high quality databases are easier to identify and access. One such example is the ZonMw program "Access to Data" (Nederlands: *Toegang tot Data*), which catalogues accessible, linkable datasets and further promotes the use of existing data sources for scientific research.[14]

Ensuring that standardized outcomes can be accurately and uniformly collected from a large number of studies is also of interest to the broader scientific community. Approximately a quarter of a trillion US dollars are allocated to biomedical research projects worldwide every year[15]; disconcertingly, a substantial proportion of this research is duplicative or may lack sufficient sample size for useful interpretation. In the recent *Lancet* series, "Research: increasing value, reducing waste", several recommendations were made to address the current level of waste in biomedical research.[16] One key suggestion was that researchers should share individual-level data.[16] To date, data sharing in public health has lagged far behind other scientific disciplines.[17] Yet, indisputably, data sharing is a cost-effective way to reduce the creation of new datasets collecting equivalent measures.[18] Moreover, data sharing has the potential to stimulate positive changes in public health policy by broadening the scope of the existing datasets.[17,19] Methodologically, pooled statistical estimates based on individual data rather than aggregate results provide greater insight into the data and allow for more sophisticated analyses.[20] The routine sharing of individual-level data further encourages researchers to undertake quality data management, which includes the creation of clean, well-annotated dataset readily interpretable by external users.[17] Documentation on cleaning procedures and additional meta-data describing the conditions of the dataset lead to greater transparency in research and allow for long-term use of data.[21]
Undoubtedly, there is sweeping support for standardization of outcomes, data sharing, and the creation of a centralized public data repository. However, as a novel endeavor in data sharing, TOPICS-MDS faces risks in data quality. For any new dataset, ensuring that data validity is upheld is fundamental. Although TOPICS-MDS contains instruments previously validated for use in older populations, this does not guarantee that these instruments produced valid results in this setting. Well-established instruments are often erroneously referred to as valid instruments. However, validity is not a property of an instrument, but rather a contextual aspect related to the study population.[22] Although greater confidence is given to instruments which have consistently produced valid results, this does not diminish the importance and necessity of re-evaluating validity. Rather, good scientific practice supports preliminary

validation work to ensure subsequent analyses are based on credible data. The presumption of data accuracy without thorough investigation can lead to faulty analyses and obscure true effects. As aptly stated by the nineteenth century French historian and political thinker, Alexis de Tocqueville,

> *…when statistics are not based upon computations which are strictly accurate, they mislead instead of guiding aright. The mind is easily imposed upon by the false affectation of exactness, which prevails even in the misstatements of science, and it adopts with confidence errors which are dressed in the forms of mathematical truth.*[23]

Public health interventions and health policy must be based on accurate data to function properly. By supporting the integrity of the data collected in TOPICS-MDS, this thesis aligns with this aim and further lays the foundation for widespread implementation of the MDS in the Netherlands. Moreover, this work gives researchers funded under the National Care for the Elderly Programme the confidence to explore health issues in a broader population, not just within their individual projects. This larger scope allows for more detailed and sophisticated analyses, thus heightening health policy impact. As an example, informal caregiver datasets are typically quite small, consisting of several hundred participants at most. TOPICS-MDS, however, has amassed data on approximately 9,000 informal caregivers and validated the use of the CarerQol, an instrument measuring care-related burden in the database. Thus, researchers have the opportunity to investigate subgroups of vulnerable caregivers which would not have been feasible in prior studies.

5.3. Limitations of TOPICS-MDS

As with any data collection tool, TOPICS-MDS is not without limitations. TOPICS-MDS is a standardized questionnaire which may not offer the level of desired specificity for some research questions. Other additional instruments as well as more detailed information on morbidity status could have been included in the dataset. Nonetheless, the use of any minimum dataset is a compromise between complete, accurate information and feasibility of data collection. Whereas more detailed data may have provided deeper insight into specific research areas, such information was not readily obtainable from all contributors. To avoid the total restructuring of individual study protocols, outcomes in TOPICS-MDS were chosen for both relevance and practicality. In light of this limitation, it is not the intention of this thesis to suggest that

TOPICS-MDS should supersede all pre-existing surveys. TOPICS-MDS was designed to be used as either a primary or complementary survey; importantly, its primary focus was to ensure that critical core outcomes are uniformly collected. This flexible design feature can, in turn, be viewed as a major strength of the survey.

In this regard, the operationalization of frailty should also be noted. The frailty index derived from TOPICS-MDS may be one of the most contentious variables within the database. To date, there is no universal consensus on the definition or identification of frailty. Some definitions refer solely to physical vulnerability (e.g. Fried's phenotype model based on five physiological indicators[24]) whereas others apply a broader interpretation including both physiological and general wellbeing attributes (e.g. the Tilburg Frailty Indicator[25] and Rockwood's frailty index[10]). Due to the structure of TOPICS-MDS, the frailty index, which defines frailty as an accumulation of health problems ("deficits") to derive an index score, was used in this thesis. However, irrespective of the definition used to identify frailty, it is important to note that there is substantial overlap and statistical convergence between these different paradigms.[26] Whereas some may not agree with the use of the frailty index, to date, there is no strong evidence that it represents an inferior operationalization of frailty. Still, given that it is conceptually difficult to define, there remains the need to better understand how to best frame and apply the frailty index in clinical practice.

Secondly, there is notable heterogeneity between individual projects included in TOPICS-MDS. Given the differences in study designs and targeted sampling populations, the unfettered analysis of the pooled dataset is not recommended. Rather, meta-analytical techniques should be applied to determine the appropriateness of combining data before performing an analysis. Based on the research question, subgroup or sensitivity analyses may also need to be built into the analysis plan. For instance, this thesis has found that differential reporting is present across survey modes when measuring activities of daily living and care-related quality of life. Sensitivity analyses by survey mode may be warranted. Although meta-analytical techniques are required, this is not to imply that meta-analysis itself is a weakness. Rather, the weakness lies in the fact that such techniques are not always in the statistical repertoire of researchers. For this reason, users may be deterred from the database or apply suboptimal analyses.

Furthermore, the pooled dataset is not representative of the general population aged 65 and older in the Netherlands. Sampling frameworks varied across individual studies included in TOPICS-MDS; some were based on a random sample, though many targeted vulnerable or disease-specific subpopulations.

Although distributions of gender (60% women), marital status (72% married) and institutionalization (9% residing in a nursing home) observed in TOPICS-MDS broadly reflect the Dutch general population aged 65 years and older (56%, 75% and 4% respectively[27]), this does not imply that the dataset is nationally representative. Differences across other key sociodemographic characteristics are likely. Design weights were not included in the development phase of TOPICS-MDS, and therefore post hoc weighting would prove highly challenging if not impracticable. However, whereas representativeness is crucial for descriptive studies, this is not necessarily true when examining causal mechanisms.[28] Arguably, for this latter type of research, greater emphasis should be placed on identifying and controlling for confounding variables.[28] Thus, despite the over-representation of certain subpopulations, TOPICS-MDS still serves as a rich resource for scientific inquiry.

Lastly, this thesis supported the validity of the data collected in TOPICS-MDS with the exception of the 15-item ADL/IADL scale. This particular scale lacked content validity for the community-dwelling older persons and was therefore unable to adequately discriminate between higher functioning persons. Future versions of TOPICS-MDS should revise this domain. Several options are available to improve physical function assessment. One option would be to adopt a more detailed scale. This does not necessarily imply that more items are required; more detailed response options (e.g. a 3- or 4-level response option rather than the current binary option) may elicit the appropriate level of discrimination. Secondly, physical measurements, such as 24-hour activity diaries, gait speed, and grip strength, can provide an objective measure of physical function suitable in both community-dwelling and institutionalized settings. However, their accuracy can be biased by the assessor's encouragement level and frequency of repetition.[29] Furthermore, compared to a survey, physical measurements are more labor intensive and costly. A third alternative is the use of item banks to identify relevant physical function measures. Ideally, an item bank would be coupled with computerized adaptive testing, a method which successively selects questions based on previous response patterns.[30] By applying a testing algorithm to determine the minimum number of relevant questions, computerized adaptive testing reduces floor and ceiling effects and quickens assessment time.[30] Similarly to physical measurements, however, this method can be cost prohibitive.

5.4. Future directions

TOPICS-MDS contains core outcomes which accurately reflect the health and wellbeing of older persons in the Netherlands. Additionally derived variables from the database, such the frailty index and TOPICS-CEP, further demonstrate its versatility. Ideally, TOPICS-MDS will continue to grow and promote standardized data collection in both academic and governmental spheres. Yet, the sustainability of the database is intrinsically linked to its ability to stimulate scientifically and socially relevant research that elucidate the changing needs of an aging population. Thus, new projects or research studies targeting (frail) older persons should consider adopting TOPICS-MDS as a base questionnaire and contributing to the central database. This would not only lead to a diverse and rich database, but also allow for direct comparative analyses between past and present research projects.

National research funding bodies, government agencies, disease registries, and medical associations serve as critical alliances in the promotion and large-scale utilization of TOPICS-MDS. Funding bodies may wish to sponsor research using the current database or mandate that TOPICS-MDS is included in new research protocols. ZonMw serves as a prime example in this regard. The organization has already sponsored two waves of TOPICS-MDS microgrant funding calls. These microgrant opportunities provided up to six months funding for early to mid-career researchers to perform investigative research using baseline and follow-up data. Furthermore, ZonMw has already incorporated TOPICS-MDS into projects funded under its dementia research and innovation program, "Memorable", which was launched in 2014. Objectives of this program include improving the level of care, support, and quality of life for persons with dementia as well as their caregivers.[31] Using TOPICS-MDS as basis for data collection, additional data elements can be added to better assess the welfare of this highly vulnerable subgroup.

On a governmental level, regularly administering TOPICS-MDS to a nationally representative sample of frail older persons would allow for longitudinal monitoring of these key outcomes. Whereas initiating a new monitoring system is cumbersome, TOPICS-MDS could potentially be incorporated into the *Ouderenmonitor.* This latter survey is carried out by the Dutch Municipal Health Service (GGD) every four years to assess the physical, mental, and social health of community-dwelling older persons.[32] A merger between these two surveys would synchronize health outcomes in research studies and public health monitoring, which is clearly advantageous.

Furthermore, TOPICS-MDS could stimulate research innovation for disease-specific registries, such as the Dutch Cancer Registry (IKNL). Registries

are still inherently single-disease focused and aim to capture disease characteristics rather than characteristics of the person with the disease. Given that cancer in older persons is often coupled with chronic co-morbidity, this narrow perspective is sub-optimal for evaluating the true extent of disease burden. Moreover, in very old persons with limited life expectancy, quality of life, physical functioning, and social wellness give a more valuable insight into successful interventions than simply measuring morbidity status or mortality. The Dutch Cancer Society (KWF) is striving to improve general quality of life in older persons with cancer through its "Cancer in the Elderly" campaign (Nederlands: *Kanker en Ouderen*).[33] The outcomes collected in TOPICS-MDS are clearly in align with this objective and would therefore complement routinely collected registry data.

Lastly, the support of medical societies plays an important role in the sustainability of the database. The Dutch Society for Clinical Geriatrics (NVKG) has recently adopted TOPICS-MDS to monitor patient wellbeing. TOPICS-MDS will have the dual purpose of collecting pertinent research data as well as assisting clinical evaluation by providing the patient's perspective on quality of care. The embracement of TOPICS-MDS by the Dutch Society for Clinical Geriatrics reflects the growing international interest in using Patient Reported Outcome Measures (PROMs) to measure quality of care. Historically, successful treatments have been defined using clinician-oriented measures, typically the presence or absence of disease. However, the morbidity tide has turned. In contemporary medicine, older persons are no longer presenting with isolated illnesses but rather a pallet of chronic conditions.[2] As health professionals adjust to meet the demands of this health transition, they often regress into the trap of dichotomized perceptions of health versus disease. However, this mindset is not befitting in an era of chronic multimorbidity and frailty which demands that older persons adapt to an increasing number of health problems that may present either slowly or rapidly. Treatment based on this old-fashioned, linear thinking does not always translate into quality of care or quality of life for older patients.

In contrast, PROMs are rooted in the perspective of the patient. By taking a patient-focused approach, PROMs can bridge communication gaps between clinicians and patients to help redefine wellbeing in the context of the individual and help jointly define successful treatment plans.[34,35] Yet, PROMs are not restricted to clinician-patient interactions. Health insurers have also acknowledged the utility of PROMs when assessing care management. PROMs add insight into patient quality of life that cannot be readily extracted from electronic patient records.[36] For these reasons, investigating how TOPICS-MDS performs as a PROM is highly relevant for the health care sector. To facilitate the use of

TOPICS-MDS as a PROM, a composite endpoint for general wellbeing (TOPICS-CEP) was developed and validated. Notably, when deriving preference-weights for domains included in TOPICS-CEP, there was a clear disparity between older persons and clinicians[37], underscoring why patient-focused measures are critical. To build on this previous work, additional research to examine the responsivity of TOPICS-MDS as a PROM is currently underway in cooperation with ZonMw. This initiative would potentially allow TOPICS-MDS to be utilized as a quality evaluation instrument as well. Arguably, this endeavor is the next step for transitioning TOPICS-MDS from a research database to a dynamic tool for integrated care.

5.5. Final remarks

From its inception, TOPICS-MDS has fostered a cooperative atmosphere between researchers across the Netherlands to create this valuable resource for evaluating the health and wellness of older persons and informal caregivers. This thesis highlights the success of this initiative and presents the first evidence that as a large pooled dataset, TOPICS-MDS contains valid and reliable information on more than 44,000 older persons and 9,000 caregivers. The collaborative approach in selecting core outcomes, its transparent infrastructure, and the accuracy of the data position TOPICS-MDS as a unique data sharing model to be emulated both nationally and internationally.

References

1. Kodner DL. The quest for integrated systems of care for frail older persons. *Aging Clin Exp Res.* Aug 2002;14(4):307-313.
2. Banerjee S. Multimorbidity--older adults need health care that can count past one. *Lancet.* Feb 14 2015;385(9968):587-589.
3. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet.* Mar 2 2013;381(9868):752-762.
4. Kodner DL, Spreeuwenberg C. Integrated care: meaning, logic, applications, and implications--a discussion paper. *Int J Integr Care.* 2002;2:e12.
5. Tsiachristas A, Hipple-Walters B, Lemmens KM, Nieboer AP, Rutten-van Molken MP. Towards integrated care for chronic conditions: Dutch policy developments to overcome the (financial) barriers. *Health Policy.* Jul 2011;101(2):122-132.
6. European Commission Innovation Union. (A Europe 2020 Initiative). Action Group B3: Integrated care. Retrieved from http://ec.europa.eu/research/innovation-union/index_en.cfm?section=active-healthy-ageing&pg=action_group_b3. Accessed November 8, 2015.
7. Theou O, Rockwood MR, Mitnitski A, Rockwood K. Disability and co-morbidity in relation to frailty: how much do they overlap? *Arch Gerontol Geriatr.* Sep-Oct 2012;55(2):e1-8.
8. Ensrud KE, Ewing SK, Cawthon PM, et al. A comparison of frailty indexes for the prediction of falls, disability, fractures, and mortality in older men. *J Am Geriatr Soc.* Mar 2009;57(3):492-498.
9. Partridge JS, Harari D, Dhesi JK. Frailty in the older surgical patient: a review. *Age Ageing.* Mar 2012;41(2):142-147.
10. Mitnitski AB, Mogilner AJ, MacKnight C, Rockwood K. The mortality rate as a function of accumulated deficits in a frailty index. *Mech Ageing Dev.* 2002;123(11):1457-1460.
11. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr.* 2008;8:24.
12. Hofman CS, Makai P, Boter H, et al. Establishing a composite endpoint for measuring the effectiveness of geriatric interventions based on older persons' and informal caregivers' preference weights: a vignette study. *BMC Geriatr.* 2014;14:51.
13. Ministerie van Volksgezondheid, Welzijn en Sport (E.I. Schippers). Brief aan de Voorzitter van de Tweede Kamer der Staten-Generaal. Een duurzaam informatiestelsel voor de zorg. Den Haag: Ministerie van VWS; 2014.
14. ZonMw. Toegang tot data. Retrieved from http://www.zonmw.nl/nl/programmas/programma-detail/toegang-tot-data-ttdata/algemeen/. Accessed November 8, 2015.
15. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet.* Jan 11 2014;383(9912):156-165.
16. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet.* Jan 18 2014;383(9913):257-266.
17. Walport M, Brest P. Sharing research data to improve public health. *Lancet.* Feb 12 2011;377(9765):537-539.
18. Wellcome Trust. Public health research data forum. Retrieved from http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030689.htm. Accessed November 8, 2015.
19. Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *Lancet.* May 14 2011;377(9778):1633-1635.
20. Stukel TA, Demidenko E, Dykes J, Karagas MR. Two-stage methods for the analysis of pooled data. *Stat Med.* Jul 30 2001;20(14):2115-2130.
21. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. *PloS One.* 2011;6(6):e21101.
22. Streiner DL, Norman GR. *Health measurement scales: a pratical guide to their development and use.* 4th ed. New York: Oxford University Press; 2008:167-210; 247-276.
23. Tocqueville AD, Bender T. *Democracy in America.* New York: Modern Library; 1981.

24. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Medical Sci*. Mar 2001;56(3):M146-156.
25. Gobbens RJ, van Assen MA, Luijkx KG, Wijnen-Sponselee MT, Schols JM. The Tilburg Frailty Indicator: psychometric properties. *J Am Med Dir Assoc*. Jun 2010;11(5):344-355.
26. Theou O, Brothers TD, Pena FG, Mitnitski A, Rockwood K. Identifying Common Characteristics of Frailty Across Seven Scales. *J Am Geriatr Soc*. Apr 2 2014.
27. Giesbers H, Verweij A, Beer J de. Vergrijzing: Wat is de huidige situatie? In: Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid. Bilthoven: RIVM; 2013. Retrieved from http://www.nationaalkompas.nl/bevolking/vergrijzing/huidig/. Accessed November 8, 2015.
28. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. Aug 2013;42(4):1012-1014.
29. Guralnick JM, Branch LG, Cummings SR, JD. C. Physical performance measures in aging research. *J Gerontol: Med Sci*. 1989;44(5):M141-146.
30. Wainer H. *Computerized Adaptive Testing*. 2nd ed. London and New York: Routledge; 2014.
31. ZonMw. Programma Memorable. Retrieved from http://www.zonmw.nl/nl/programmas/programma-detail/memorabel/algemeen/. Accessed November 8, 2015.
32. GGD Gelderland-Zuid. Ouderenmonitor. Retrieved from https://www.ggdgelderlandzuid.nl/nl-NL/Professionals/Gemeenten/Gezonde%20wijk/Themas-en-instrumenten/Ouderenmonitor.aspx. Accessed November 8, 2015.
33. KWF Kanker Bestrijding. Focus 2011-2014: Kanker en ouderen. Retrieved from https://www.kwf.nl/over-kwf/Pages/Kanker-en-ouderen.aspx. Accessed November 8, 2015.
34. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindbla S. Patient reported outcome measures in practice. *BMJ*. 2015;350:g7818
35. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-68.
36. Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood)*. Oct 2010;29(10):1863-1871.
37. Hofman CS, Makai P, Blom JW, et al. Comparing the health state preferences of older persons, informal caregivers and healthcare professionals: a vignette study. *PloS One*. 2015;10(3):e0119197.

6

Nederlandse Samenvatting

Belangrijkste bevindingen

- De 15-item ADL/IADL-schaal heeft weinig onderscheidend vermogen in hoger functionerende ouderen. In toekomstige versies van de MDS zal het domein 'fysiek functioneren' moeten worden herzien om dit segment van de bevolking beter te representeren.
- *Survey mode* (manier van uitvragen) beïnvloedt rapportage van urine-incontinentie, lopen en maaltijdbereiding in residentiële zorginstellingen.
- Daar waar de EQ-5D, een instrument dat gezondheidsgerelateerde kwaliteit van leven meet, een plafondeffect laat zien in de totale onderzoekspopulatie, is dit niet het geval binnen subgroepen met belangrijke geriatrische aandoeningen.
- De validiteit van de CarerQol, een meetinstrument voor zorggerelateerde kwaliteit van leven, wordt gehandhaafd ongeacht het steekproefkader (uit algemene bevolking, eerstelijn, of ziekenhuissetting), hoewel er verschillen werden waargenomen voor *survey mode*.
- Een kwetsbaarheidsindex die afgeleid is van TOPICS-MDS is een valide maat voor kwetsbaarheid.
- Bij de beoordeling van kwaliteit van leven en kosten voor de gezondheidszorg, is kwetsbaarheid een betekenisvolle onafhankelijke covariaat, naast multimorbiditeit en ADL-beperkingen.
- Een samengestelde maat, afgeleid van TOPICS-MDS en gebaseerd op gewogen preferenties, vormt een geldige indicatie van gezondheidsgerelateerde kwaliteit van leven die nuttig is voor evaluatiestudies.

6.1 Samenvatting van bevindingen

Hoewel het delen van gegevens op veel steun kan rekenen van de theoretische wetenschappelijke gemeenschap, wordt het tot op heden nog niet op grote schaal in de praktijk toegepast. Dit is ten dele te wijten aan een gebrek aan samenwerking tussen onderzoekers. TOPICS-MDS is in die zin een baanbrekend initiatief gebleken, aangezien de ontwikkeling ervan niet alleen de inzet vereiste van de acht Nederlandse universitaire medische centra, maar ook van tal van individuele onderzoekers in heel Nederland. Alle hoofdonderzoekers, wiens onderzoek in aanmerking kwam voor subsidie in het Nationaal Programma Ouderenzorg, zijn bij aanvang overeengekomen om TOPICS-MDS te integreren in hun studieprotocollen. Het product hiervan is een centraal samengevoegd publiek toegankelijke database met gestandaardiseerde gezondheidsuitkomsten die verzameld zijn uit meer dan 60 verschillende onderzoeksprojecten. De ontwikkeling van de TOPICS-MDS database vormt zonder twijfel een belangrijke mijlpaal in het delen van gegevens en kent vele mogelijkheden voor onderzoek. Toch dient elk nieuw instrument voor gegevensverzameling te worden benaderd met enige omzichtigheid totdat de nauwkeurigheid en kwaliteit van de gegevens op passende wijze zijn geëvalueerd.

De overkoepelende doelstelling van het promotieonderzoek beschreven in deze dissertatie was om de gegevens verzameld in TOPICS-MDS te valideren. Om dit doel te bereiken werden in Hoofdstuk 3 de meeteigenschappen beoordeeld van de drie primaire instrumenten in TOPICS-MDS: de Katz Index of Independence in Activities of Daily Living (Katz-ADL index), de EQ-5D en de CarerQol. In Hoofdstuk 4 werden twee afgeleide variabelen geëvalueerd die kwetsbaarheid en gezondheidsgerelateerde kwaliteit van leven meten. Aangezien kwetsbaarheid een belangrijke indicator is in geriatrisch onderzoek, werden in paragrafen 4.1 en 4.2 een afgeleide maat van de kwetsbaarheidsindex (d.w.z. een samengestelde maat gebaseerd op een spectrum van gezondheidsproblemen) gevalideerd met behulp van TOPICS-MDS data en werd de toegevoegde waarde van het meten van kwetsbaarheid in relatie tot multimorbiditeit en activiteiten van het dagelijks leven onderzocht. Gezien de groeiende belangstelling voor meer alomvattende indicatoren voor welzijn, werd in paragraaf 4.3 een samengestelde maat voor gezondheidgerelateerde kwaliteit van leven gevalideerd (TOPICS-CEP).

Hoofdstuk 3, paragrafen 3.1 en 3.2 zijn gericht op de gewijzigde Katz-ADL index, die onderzoekt of ouderen hulp nodig hebben bij 15 basale en instrumentele activiteiten. In paragraaf 3.1 werd een Rasch-analyse toegepast om te bepalen of de moeilijkheidsgraad van deze 15 items op dezelfde manier werd beoordeeld door ouderen die thuiswonend zijn als door ouderen die in residentiële zorginstellingen verblijven. Daarnaast werd onderzocht of de schaal een robuust onderscheidend vermogen heeft. Deze analyse trof kleine verschillen aan in de waardering van de moeilijkheidsgraad, waardoor een directe vergelijkende analyse bemoeilijkt wordt. Bij dergelijke analyses, kan toepassing van gewogen items nodig zijn om moeilijkheidsgraden van items tussen deze twee settings te standaardiseren. Daarnaast blijkt de schaal onvoldoende discriminerend vermogen te laten zien voor hoger functionerende thuiswonende ouderen. In toekomstige versies van TOPICS-MDS, zal het domein 'fysiek functioneren' moeten worden herzien om dit segment van de bevolking beter te representeren. In een aanvullende analyse in paragraaf 3.2, werd een Rasch-analyse toegepast om te onderzoeken of *survey mode* (d.w.z. schriftelijke vragenlijst versus interview) leidde tot differentiële rapportage van activiteiten in residentiële zorginstellingen. *Survey mode* bleek rapportage van urine-incontinentie, lopen en maaltijdbereiding te beïnvloeden, waardoor het toepassen van weging of interactietermen nodig kan zijn bij het analyseren van een samengevoegde *mixed-mode* dataverzameling.

In Hoofdstuk 3, paragraaf 3.3 werden de meeteigenschappen van de EQ-5D, een instrument voor gezondheidsgerelateerde kwaliteit van leven, voor vier belangrijke geriatrische aandoeningen beoordeeld: gehoorstoornissen, gewrichtsschade, urine-incontinentie en duizeligheid met vallen. Resultaten werden vergeleken met de gehele studiepopulatie en een "gezonde" subgroep zonder chronische aandoeningen. Onafhankelijk van classificatie van morbiditeit, correleerden individuele domeinen van de EQ-5D goed met andere vragenlijstitems die vergelijkbare concepten meten en volgden ze de *a priori* gestelde hypothesen over associaties met sociodemografische factoren, waardoor constructvaliditeit wordt ondersteund. Als gevolg van plafondeffecten vertoonde de EQ-5D weinig onderscheidend vermogen voor ouderen die over het algemeen gezond waren, dat wil zeggen, die weinig tot geen chronische aandoeningen rapporteerden. Hoewel de EQ-5D aanvankelijk een slecht onderscheidend vermogen liet zien binnen de totale dataset, was dat niet het geval voor subgroepen met belangrijke geriatrische aandoeningen.

In Hoofdstuk 3, paragraaf 3.4 werd het Care-Related Quality of Life Instrument (CarerQol) gevalideerd met betrekking tot twee verschillende kenmerken van het studiedesign. Namelijk het steekproefkader (algemene bevolking versus diverse zorginstellingen) en de *survey mode* (interview versus

schriftelijke vragenlijst). De waargenomen correlaties tussen de CarerQol-domeinen en een generieke maat voor geluk en zelfgerapporteerde belasting ondersteunde de constructvaliditeit van dit instrument. Op basis van multivariabele, multivariate mixed models, had het steekproefkader slechts een geringe invloed op de rapportageniveaus. Echter, *survey mode* leek een bescheiden invloed te hebben op zelfgerapporteerde belasting. Toekomstige analyses dienen rekening te houden met de impact van de manier van uitvragen op de rapportage van zorggerelateerde kwaliteit van leven; het gebruik van een interactieterm in multivariabele modellen zou een optie kunnen zijn om dit probleem aan te pakken.

In Hoofdstuk 4, paragrafen 4.1 en 4.2 werd gekeken naar de validiteit en de operationalisering van de kwetsbaarheidsindex. In paragraaf 4.1 werden een lange (46-item) en een korte (23-item) kwetsbaarheidsindex, gebaseerd op TOPICS-MDS, gecorreleerd met een kwetsbaarheidsindex op basis van een klinische beoordeling. De lange en korte TOPICS kwetsbaarheidsindexen bleken matig te correleren met de onafhankelijke klinisch geëvalueerde kwetsbaarheidsindex ("referentiestandaard"). Bovendien volgden beide indexen de *a priori* gestelde hypothesen met betrekking tot distributie van leeftijd en geslacht, waardoor de validiteit van beide indexen ondersteund wordt. In paragraaf 4.2 werd ingegaan op de operationalisering van kwetsbaarheid en de relatie ervan met multimorbiditeit en ADL-beperkingen. Kwetsbaarheid, multimorbiditeit en ADL-beperkingen werden gerapporteerd voor een aanzienlijk deel van de ouderen in TOPICS-MDS. Bovendien rapporteerden ouderen die problemen ondervinden in deze drie domeinen, de laagste scores op kwaliteit van leven en de hoogste kosten voor gezondheidszorg. Op basis van multivariabele mixed models die scores op kwaliteit van leven en kosten voor de gezondheidszorg meten, bleek kwetsbaarheid een belangrijke onafhankelijke covariaat, naast multimorbiditeit en ADL-beperkingen.

In hoofdstuk 4, paragraaf 4.3 werd een gecombineerde uitkomstmaat voor de door de oudere ervaren gezondheidsgerelateerde kwaliteit van leven gevalideerd. Deze Composite End Point (TOPICS-CEP) werd berekend op basis van prioritering van uitkomstmaten in TOPICS-MDS door ouderen en mantelzorgers. Op basis van de meta-correlaties, bleek TOPICS-CEP matig gecorreleerd te zijn met Cantril's Self-Anchoring Ladder (een maat voor algemene kwaliteit van leven) en sterk gecorreleerd te zijn met de EQ-5D (een maat voor gezondheidsgerelateerde kwaliteit van leven). Op basis van multivariabele mixed models, bleken TOPICS-CEP scores hoger voor personen die getrouwd waren, zelfstandig woonden, of een universitaire opleiding hadden genoten, en lager voor ouderen met dementie, depressie, of duizeligheid met vallen. De TOPICS-CEP volgde de *a priori* gestelde hypotheses en ondersteunt dus de construct en bekende-groep

validiteit (*known-group validity*). Deze studie onderbouwt het gebruik van TOPICS-CEP als een betekenisvolle indicator van de gezondheidgerelateerde kwaliteit van leven in evaluatieonderzoek.

7

Supplement

7.1. Acknowledgements

I would like to thank my promoter, Professor Marcel Olde-Rikkert, for sharing with me his wisdom, guidance, and unwavering support for women in science. I would also like to thank my co-promoter, Dr. René Melis, for his keen analytical insight and engaging epidemiological dialogue. I extend my deepest gratitude to all TOPICS consortium members, past and present, for their contribution to this thesis. Very special recognition goes out to former TOPICS project manager and translator extraordinaire, Dr. Dannie van den Brink, whose professional and personal support helped me stay focused on what truly matters. Many thanks to my partner in crime, Cynthia Hofman, for her friendship and abundance of good ideas – Believe and Achieve! I reserve special thanks for former TOPICS working group members, Dr. Lia Baars and Dr. Bianca Schalk, as well as for my mentor, Dr. Nel Roelveld, who were always available to give me a healthy serving of sage advice peppered with inspiration. I would like to acknowledge Dr. Rogier Donders for his cheerful statistical guidance and witticisms. I am grateful for the support from my colleagues in the Department of Geriatric Medicine, in particular my roomies Noralie Geessink, Maaike Pijnenborg, Anne Rijmpa, Anke Richters, and Anouk Vermeij. I would be remiss not to acknowledge all the NPO contributors as well as the participants who volunteered their time and personal health data – without them this project would not have been possible. I would also like to recognize the educators at SUNYA SPH for giving me the initial toolset to pursue this goal, in particular Mike Zdeb, whose lifetime guarantee was legit! To my crazy Lutomski/Tremacco family, thanks for all the wonderful and weird times we have shared together - love you guys to the moon and back! Last but not least, I must thank my amazing husband, Jaap Verheggen, my North Star, ever present to guide me through times of darkness, my beautiful daughter, Ilona, who re-affirms my faith in miracles every day, and Lee and Missy, for all the much needed furry therapy.

Set your sights upon the heights.
Don't be a mediocrity.
Don't just wait and trust to fate,
and say that's how it's meant to be.
It's up to you how far you go.
If you don't try, you'll never know.
And so my lad as I've explained,
nothing ventured, nothing gained.

-Merlin *(The Sword in the Stone,* 1963*)*

7.2. Peer-reviewed publications

Thesis publications

1. **Lutomski JE**, Baars MA, Schalk BW, Boter H, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis RJ; TOPICS-MDS Consortium. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. *PLoS One*. 2013 Dec 4;8(12):e81673. doi: 10.1371/journal.pone.0081673.
2. **Lutomski JE**, Krabbe PFM, Bleijenberg N, Blom JW, Buurman BM, Muntinga ME, Steyerberg EW, Olde Rikkert MGM, Kempen GIJM, Melis RJF. Rasch analysis reveals combining ADL/IADL summary scores from different residential settings is inappropriate. *J Clin Epi*. (In press)
3. **Lutomski JE**, Krabbe PFM, Bleijenberg N, Blom JW, Buurman BM, Kempen GIJM, Muntinga ME, Steyerberg EW, Olde Rikkert MGM, Melis RJF. ADL and IADL reporting are biased by survey mode. *J Am Geriatr Soc*. 2015 Nov;63(11):2419-21.
4. **Lutomski JE**, Krabbe PFM, Bleijenberg N, Blom JW, Kempen GIJM, MacNeil-Vroomen J, Muntinga ME, Steyerberg EW, Olde Rikkert MGM, Melis RJF. Measurement properties of the EQ-5D across four major geriatric conditions: Findings from TOPICS-MDS. (Under review)
5. **Lutomski JE**, van Exel NJ, Kempen GI, Moll van Charante EP, den Elzen WP, Jansen AP, Krabbe PF, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis RJ. Validation of the Care-Related Quality of Life Instrument in different study settings: findings from The Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS). *Qual Life Res*. 2015 May;24(5):1281-93.
6. **Lutomski JE**, Baars MA, van Kempen JA, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Krabbe PF, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis RJ. Validation of a frailty index from the older persons and informal caregivers survey minimum data set. *J Am Geriatr Soc*. 2013 Sep;61(9):1625-7.
7. **Lutomski JE**, Baars MA, Boter H, Buurman BM, den Elzen WP, Jansen AP, Kempen GI, Steunenberg B, Steyerberg EW, Olde Rikkert MG, Melis R. [Frailty, disability and multi-morbidity: the relationship with quality of life and healthcare costs in elderly people]. *Ned Tijdschr Geneeskd*. 2014;158:A7297. In Dutch.
8. Hofman C, **Lutomski JE**, Boter H, Buurman BM, de Craen A, Donders RA, Olde Rikkert MGM, Makai P, Melis RJF. Examining the construct and

known-group validity of a composite endpoint for The Older Persons and Informal Caregivers Survey Minimum Data Set (TOPICS-MDS). (Under review)

9. van den Brink D, **Lutomski JE**, Qin L, den Elzen WP, Kempen GI, Krabbe PF, Steyerberg EW, Muntinga M, Moll van Charante EP, Bleijenberg N, Olde Rikkert MG, Melis RJ; namens het TOPICS-MDS Consortium. [TOPICS-MDS: a versatile resource for generating scientific and social knowledge for elderly care]. *Tijdschr Gerontol Geriatr*. 2015 Apr;46(2):78-91. In Dutch.

Other publications - Accepted/published

10. Hofman C, **Lutomski JE**, Olde Rikkert MGM, Melis RJ, Donders AR. Examining consistency between self-assessed QOL and peer ratings in older people. *J Am Geriatr Soc*. (In press)
11. Schepens MEW, **Lutomski JE**, Bruce I, Olde Rikkert MGM, Lawlor BA. The reliability and validity of the International Dementia Alliance (IDEAL) schedule for the assessment and staging of care in Ireland. *Am J Geriatr Psychiatry* (In press)
12. **Lutomski JE**, Meaney S, Greene RA, Ryan AC, Devane D. Expert systems for fetal assessment in labour. *Cochrane Database Syst Rev*. 2015 Apr 30;4: CD010708. doi: 10.1002/14651858.CD010708.pub2.
13. **Lutomski JE**, Donders AR, Melis RJ. Causal diagrams to better understand missingness. *JAMA Pediatr*. 2014 Feb;168(2):187.
14. Meaney S, Gallagher S, **Lutomski JE**, O'Donoghue K. Parental decision making around perinatal autopsy: a qualitative investigation. *Health Expect*. 2014 Nov 6. doi: 10.1111/hex.12305.
15. McCarthy FP, **Lutomski JE**, Greene RA. Hyperemesis gravidarum: current perspectives. *Int J Women's Health*. 2014 Aug 5;6:719-25. doi: 10.2147/IJWH.S37685.
16. **Lutomski JE**, Murphy M, Devane D, Meaney S, Greene RA. Private health care coverage and increased risk of obstetric intervention. *BMC Pregnancy Childbirth*. 2014 Jan 13;14:13. doi: 10.1186/1471-2393-14-13.
17. **Lutomski JE**, Shiely F, Molloy EJ. The prevalence of syphilis at childbirth in Ireland: a six-year review. *J Matern Fetal Neonatal Med*. 2014 Nov;27(17):1823-5.
18. Khalid AS, Marchocki Z, Hayes K, **Lutomski JE**, Joyce C, Stapleton M, O'Mullane J, O'Donoghue K. Establishing trimester-specific maternal thyroid function reference intervals. *Ann Clin Biochem*. 2014 Mar;51(Pt 2):277-83.
19. O'Neill SM, Kearney PM, Kenny LC, Henriksen TB, **Lutomski JE**, Greene RA, Khashan AS. Caesarean delivery and subsequent pregnancy interval: a systematic review and meta-analysis. *BMC Pregnancy Childbirth*. 2013 Aug 27;13:165. doi: 10.1186/1471-2393-13-165.

20. Upson K, Silver RM, Greene R, **Lutomski J**, Holt VL. Placenta accreta and maternal morbidity in the Republic of Ireland, 2005-2010. *J Matern Fetal Neonatal Med*. 2014 Jan;27(1):24-9.
21. Fawsitt CG, Bourke J, Greene RA, Everard CM, Murphy A, **Lutomski JE**. At what price? A cost-effectiveness analysis comparing trial of labour after previous caesarean versus elective repeat caesarean delivery. *PLoS One*. 2013;8(3):e58577. doi: 10.1371/journal.pone.0058577.
22. **Lutomski J**, Dempsey E, Molloy E. An epidemiological study of factors associated with preterm infant in-hospital mortality. *Ir Med J*. 2013 Jan; 106(1):9-12.
23. O'Neill SM, Khashan AS, Kenny LC, Greene RA, Henriksen TB, **Lutomski JE**, Kearney PM. Caesarean section and subsequent ectopic pregnancy: a systematic review and meta-analysis. *BJOG*. 2013 May;120(6):671-80.
24. O'Neill SM, Kearney PM, Kenny LC, Khashan AS, Henriksen TB, **Lutomski JE**, Greene RA. Caesarean delivery and subsequent stillbirth or miscarriage: systematic review and meta-analysis. *PLoS One*. 2013;8(1):e54588. doi: 10.1371/journal.pone.0054588.
25. **Lutomski JE**, Morrison JJ, Lydon-Rochelle MT. Regional variation in obstetrical intervention for hospital birth in the Republic of Ireland, 2005-2009. *BMC Pregnancy Childbirth*. 2012 Nov 5;12:123. doi: 10.1186/1471-2393-12-123.
26. **Lutomski JE**, Greene RA, Byrne BM. Severe maternal morbidity during childbirth hospitalisation: a comparative analysis between the Republic of Ireland and Australia. *Eur J Obstet Gynecol Reprod Biol*. 2012 Aug;163(2):148-53.
27. **Lutomski JE**, Byrne BM, Devane D, Greene RA. Increasing trends in atonic postpartum haemorrhage in Ireland: An 11-year population-based cohort study. *BJOG*. 2012; 119(3):306-314.
28. **Lutomski JE**, Morrison JJ, Greene RA, Lydon-Rochelle MT. Maternal morbidity during delivery hospitalization. *Obstet Gynecol*. 2011; 117(3):569-602.
29. Browne GM, Eustace JA, Fitzgerald AP, **Lutomski JE**, Perry IJ. Prevalence of diminished kidney function in a representative sample of middle and older age adults in the Irish population. *BMC Nephrol*. 2012 Nov 2;13:144. doi: 10.1186/1471-2369-13-144.
30. Harrington J, Fitzgerald AP, Layte R, **Lutomski J**, Molcho M, Perry IJ. Socio-demographic, health and lifestyle predictors of poor diets. *Public Health Nutr*. 2011; June 13:1-10.
31. Shiely F, Perry IJ, **Lutomski J**, Harrington J, Kelleher CC, McGee H and Hayes K. Temporal trends in misclassification patterns of measured and self-report based body mass index categories. *BMC Public Health*. 2010; Sep 17; 10:560.

32. **Lutomski JE**, Van den Broeck J, Harrington J, Shiely F, and Perry IJ. Socio-demographic, lifestyle, mental health and dietary factors associated with direction of misreporting of energy intake. *Public Health Nutr.* 2010; Aug 16:1-10.
33. Harrington J, Perry I, **Lutomski J**, Fitzgerald AP, Shiely F, McGee H, Barry MM, Van Lente E, Morgan K and Shelly E. Living longer and feeling better: healthy lifestyle, self-rated health, obesity and depression in Ireland. *Eur J Public Health*. 2010; 20(1):91-5.

Other publications - Under review

34. Ignatov PN, **Lutomski JE**. Quantitative cardiotocography to improve fetal assessment during labor: A randomized controlled trial.
35. Fawsitt C, Bourke J, **Lutomski JE**, Meaney S, McElroy B, Murphy R, Greene RA. What women want: exploring pregnant women's preferences for alternative models of maternity care.
36. Fawsitt C, Bourke J, McElroy B, Murphy R, Greene RA, **Lutomski JE**. Eliciting women's preferences for alternative models of maternity care in Ireland: A discrete choice experiment.
37. Fawsitt C, Bourke J, Murphy A, McElroy B, **Lutomski JE**, Murphy R, Greene RA. A cost-benefit analysis of consultant- and midwifery-led care in Ireland: a feasibility study using the contingent valuation method.
38. Fawsitt C, Bourke J, Murphy A, McElroy B, Murphy R, **Lutomski JE**, Greene RA. Choice consistency and mode specification: An examination of the dual response choice format.
39. Meaney S, **Lutomski JE**, O' Connor L, O' Donoghue K, Greene RA. Women's experience of maternal morbidity: A qualitative analysis.
40. Meaney S, Corcoran P, **Lutomski JE**, Gallagher S, Spillane N, O'Donoghue K. Perceived maternal stress and emotional wellbeing as risk factors for miscarriage.

7.3. Biography of the author

Jennifer Elizabeth Lutomski grew up surrounded by the scenic mountains of the Hudson Valley (New York, USA). Her desire to "think global, act local" inspired her to enter the field of public health. She completed a Master of Science in Epidemiology and a Certificate of Social Demography at the University at Albany School of Public Health (Albany, New York, USA). During her studies, a chance encounter with a Dutchman (her later husband) was the initial spark which ignited a transcontinental move to Europe. Arm in arm with her husband, Jennifer moved to Cork, Ireland, where she worked for University College Cork. After two years as a research assistant in the Department of Epidemiology and Public Health, she accepted a promotion as the unit epidemiologist in the National Perinatal Epidemiology Centre (Department of Obstetrics and Gynaecology). Despite six brilliant years on the Emerald Isle, Jennifer and her husband decided it was time to move again and relocated to Wijchen, The Netherlands. Her growing interest in taking more of a leadership position in epidemiological research provoked Jennifer to pursue a PhD with the Department of Geriatric Medicine, Radboud University Medical Center (Nijmegen, The Netherlands).

Jennifer has served on two large national consortiums [2007 Survey of Lifestyle, Attitudes and Nutrition (SLÁN, Ireland); The Older Persons and Informal Caregivers Minimum Dataset (TOPICS-MDS, The Netherlands)] and has published numerous reports and scientific articles. She has worked on several national and international grants. Since she always loves the opportunity to talk about epidemiology, Jennifer has coordinated and taught bachelor-level courses in epidemiology as well as given invited lectures for various departments.

A nature-lover and avid baker, Jennifer is raising a spunky daughter with her fun-loving physicist husband. She is an advocate for animal rescue and shares her home with two retired racing greyhounds.